



Leverage, Derivatives, and Asset Markets

Citation

Yang, David Cherngchiun. 2015. Leverage, Derivatives, and Asset Markets. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:17467294>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Leverage, Derivatives, and Asset Markets

A dissertation presented

by

David Cherngchiun Yang

to

The Department of Business Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Business Economics

Harvard University

Cambridge, Massachusetts

May 2015

© 2015 David Cherngchiun Yang
All rights reserved.

Dissertation Advisors:

Professor Lawrence H. Summers

Professor John Y. Campbell

Author:

David Cherngchiun Yang

Leverage, Derivatives, and Asset Markets

Abstract

This dissertation consists of three independent essays on the relationship between leverage, derivatives (especially, option securities), and asset markets. Chapter 1, "Does the Tail Wag the Dog? How Options Affect Stock Price Dynamics," demonstrates empirically that the existence and trading of financial options affects the price movements of their underlying assets, due to the implicit leverage in options and the hedging behavior of options sellers. These empirical results contrast with classical asset pricing where options instead derive their value from their underlying assets. Chapter 2, "Disagreement and the Option Stock Volume Ratio," examines a variable known as the option stock volume ratio, which prior work has documented to be a negative predictor of stock returns in the cross section. I propose an alternate explanation based on the behavioral finance literature on belief disagreement between investors. I show how my disagreement model makes predictions in line with prior empirical findings and can also better explain other stylized facts, which I document. Chapter 3, "Bond Fire Sales and Government Interventions," analyzes how a government should intervene in response to a fire sale in the bond market. I contrast the policies of the government directly purchasing financial securities vs the government offering leverage to the private sector to purchase securities.

Contents

Abstract	iii
Acknowledgments	viii
Introduction	1
1 Does the Tail Wag the Dog? How Options Affect Stock Price Dynamics	3
1.1 Introduction	3
1.2 Theoretical Framework and Hedging Demand	9
1.2.1 Notation and Option Terminology	9
1.2.2 Hedging and Re-Hedging	10
1.2.3 Price Impact of Re-Hedging	13
1.2.4 Hedging Demand and its Empirical Counterpart	15
1.3 Dataset and Summary Statistics	17
1.3.1 Time Series and Cross Sectional Patterns in Our Dataset	18
1.4 Main Empirical Results	22
1.4.1 Baseline Regression	22
1.4.2 Robustness: Different Subsamples and Estimation Methods	27
1.4.3 Effect on Long Horizon Returns	34
1.4.4 Systematic vs Idiosyncratic Components of Returns	36
1.5 Portfolio Sorting Strategy	37
1.6 Identification via Instrumental Variable	41
1.7 Conclusion	49
2 Disagreement and the Option Stock Volume Ratio	51
2.1 Introduction	51
2.2 A Model of Disagreement and Derivatives Trade	56
2.3 Data, Variable Definitions, and Summary Statistics	62
2.4 Main Empirical Results	64
2.4.1 Correlations with O/S	67
2.4.2 Long-Short Portfolios	68
2.4.3 Effect of Changing the Portfolio Formation Period	73

2.4.4	Interaction Effects with Analyst Forecast Dispersion	78
2.5	Conclusion	81
3	Bond Fire Sales and Government Interventions	83
3.1	Introduction	83
3.2	A Model of Fire Sales	88
3.2.1	Constrained Equilibrium with Debt Overhang	90
3.3	Government Interventions	92
3.3.1	Direct Purchases	92
3.3.2	Equity Injections	93
3.3.3	Offering Leverage for Purchases	94
3.3.4	Numerical Simulation: Offering Leverage vs Direct Purchases	97
3.4	Discussion	100
3.4.1	Potential Extensions of the Model	100
3.4.2	U.S. and U.K. Government Interventions	103
3.5	Conclusion	108
	Bibliography	110
	Appendix A Proofs to Chapter 2	117
	Appendix B Proofs to Chapter 3	120

List of Tables

1.1	Closed Form Solution of Impulse Response	15
1.2	List of Key Variables	19
1.3	Summary Statistics	23
1.4	Baseline Results	26
1.5	Effect on Daily Turnover	28
1.6	Baseline Results, Using Different Subsamples of the Main Dataset	30
1.7	Baseline Results, Using Different Subsamples of Options	31
1.8	Baseline Results, Using Different Estimation Methods	32
1.9	Decomposing Returns into Systematic and Idiosyncratic Components	38
1.10	Alpha of a Long-Short Portfolio (Weekly Holding Periods)	40
1.11	Alpha per Quintile (Weekly Holding Periods)	42
1.12	IV, Distance to Nearest Round Number: 2SLS	48
2.1	List of Key Variables	65
2.2	Summary Statistics	66
2.3	Correlation Table of Quintiles of Selected Variables	67
2.4	Alpha of a Long-Short Portfolio	70
2.5	Alpha of a Long-Short Portfolio: <i>O/S</i> , Different Subsamples	71
2.6	Fama-MacBeth Regression: Cross Section of Individual Stock Returns	74
2.7	Interaction Effect with Analyst Forecast Dispersion	80

List of Figures

1.1	How Hedged Option Writing Affects Stock Prices	11
1.2	Time Series and Cross Sectional Properties of Log Hedging Demand	21
1.3	Impact at Long Horizons	35
1.4	IV, Distance to Nearest Round Number: First Stage Graphs	44
2.1	Monthly Alpha at Different Lags	77
3.1	Numerical Simulation of Direct Purchases vs Offering Leverage	99

Acknowledgments

I am grateful to my advisors John Campbell, Robin Greenwood, and Lawrence Summers for their advice and mentorship. They shaped my Ph.D. experience on everything from high-level conversations on academia as a career, to nurturing my early research projects, to reading multiple drafts of my papers. I hope I can be as good a mentor for others as they have been for me.

Conversations with other faculty members, including Malcolm Baker, Lauren Cohen, Sam Hanson, Alp Simsek, Andrei Shleifer, Erik Stafford, Jeremy Stein, and Adi Sunderam, sharpened my thinking on the essays in this dissertation as well as on other research projects I worked on during graduate school. John's, Alp's, and Jeremy's classes provided an excellent foundation in the fields of asset pricing, frictional finance, and behavioral finance. The essays in this dissertation connect closely to ideas I first learned in their courses.

Fellow graduate students as well as other members of the Harvard community enriched my time in graduate school, both in and out of the office. Many of the insightful, funny, and memorable moments during the past several years were with them. Special thanks to: Natalie Bau, Marah Brown, Kelly Friendly, Paul Goldsmith-Pinkham, Sam Kruger, Jamie Lee, Sarah Mapes, Vladimir Mukharlyamov, Charles Nathanson, Tom Powers, Natasha Sarin, Jesse Schreger, Julie Shample, Tom Wollmann, Fan Zhang (with whom I co-authored Chapter 1), John Zhou, and Eric Zwick.

The National Science Foundation's Graduate Research Fellowship Program, Harvard Business School's Doctoral Programs, and Harvard University's Dissertation Completion Fellowship provided financial support for my research. Harvard Business School's Baker Research Services and Research Computing Services provided assistance with databases and the research computing grid.

Most of all, I thank my family for their unconditional love, support, and encouragement.

Introduction

This dissertation consists of three independent essays on the relationship between leverage, derivatives (especially, option securities), and asset markets. Chapter 1 studies how the implicit leverage in option securities can affect the underlying market. Chapter 2 studies how belief disagreement can explain why options volume is a negative predictor of cross-sectional stock returns. Chapter 3 studies how the government can use leverage to alleviate the consequences of bond fire sales.

In Chapter 1, "Does the Tail Wag the Dog? How Options Affect Stock Price Dynamics" (co-authored with Fan Zhang), we demonstrate empirically that the existence and trading of options affects the autocorrelation of stock returns. Dynamic hedging by option writers creates an upward sloping demand curve: when the stock price rises, option writers must buy more stock to remain hedged. Because option writers do not adjust their hedges instantaneously, their hedging increases the autocorrelation of stock returns. We study how the hedging of options on individual stocks affects stock return autocorrelations in the cross section. We develop a measure of "hedging demand" that quantifies the sensitivity of the option writers' hedge to changes in the underlying price. As we move from the lowest to highest quintile of hedging demand, the daily return autocorrelation increases from -5.0% to -1.6%. In a portfolio sorting strategy, the high-minus-low-quintile return spread has a gross alpha of 0.33% weekly (18% annualized). To address potential confounds, we use an instrumental variable for hedging demand, the absolute difference between the underlying stock price and the nearest round number. This instrument uses the institutional idiosyncrasy that exchange-traded options have round number strike prices.

In Chapter 2, "Disagreement and the Option Stock Volume Ratio," I analyze the relationship between options volume and returns in the underlying stock. Recent research has shown that the ratio of options trading volume to stock trading volume (O/S) is a negative predictor of cross-sectional stock returns. Johnson and So (2012) document this empirical finding and explain the predictability using an asymmetric information model, in which informed traders prefer to express negative views using options. In this essay, I instead propose a disagreement model to explain the negative relationship between O/S and cross-sectional stock returns. This disagreement model makes predictions in line with prior empirical findings and can also better explain other empirical facts about O/S , including (1) the relationship of O/S with proxies for disagreement; (2) the ability of both put and call volume to forecast negative returns in the cross section; (3) the long-term predictability of returns in the cross section using O/S ; (4) interaction effects with proxies for disagreement.

In Chapter 3, "Bond Fire Sales and Government Interventions," I analyze how a government should intervene in response to a fire sale in the bond market. Fire sales affect the real economy by increasing the hurdle rate on new bond issuances and thereby dampening investment. I build a model to analyze two questions: When there is a bond fire sale, should the government intervene by directly purchasing securities or by offering leverage to the private sector for their own purchases? And, should the government intervene in the primary market for new bond issuances or in the secondary market for existing bonds? My model shows that when the government pursues direct purchases, the choice of primary vs secondary market does not matter because bonds in the two markets are perfect substitutes. However, by offering leverage only to new bonds, the government can break the perfect substitutability. Prices across the two markets can now deviate, and the government can lower the interest rate on new bonds with less "leakage" into the secondary market. Hence, to aid the real economy in response to bond fire sales, the government should focus on providing leverage for the purchase of new bonds. I also discuss how my model relates to interventions by the U.S. and U.K. governments aimed at stimulating new lending.

Chapter 1

Does the Tail Wag the Dog? How Options Affect Stock Price Dynamics¹

1.1 Introduction

Does the existence and trading of financial options influence the price movements of their underlying assets? In classical asset pricing, options instead “derive” their value from their underlying assets. In this paper, we document how options on individual stocks affect their underlying assets through the hedging behavior of option writers.

In our model, the end users are option buyers. Option writers sell options to accommodate end user demand and hedge their exposure by dynamically trading the underlying stock (Black and Scholes, 1973; Merton, 1973). Dynamic hedging by option writers creates an upward sloping demand curve: when the underlying stock price rises, option writers must buy more of the stock to remain hedged. This upward sloping demand curve applies for an option writer who has sold either call options or put options. For hedging a call option, the intuition of buying more of the underlying as the stock price rises applies naturally. For hedging a put option, it is more natural to think about the upward sloping demand curve as short selling more of the underlying as the stock price falls.

¹Co-authored with Fan Zhang

If hedging occurs at a lag and there are demand pressure effects, then the model predicts that hedging by option writers increases the autocorrelation of stock returns.² For example, suppose there is a positive news shock at period 1, which causes prices to rise. At period 2, option writers then re-adjust their hedges, which forces them to purchase more stock. This re-hedging causes prices to rise even further and hence creates autocorrelation in the stock returns. Theoretical papers in mathematical finance have explored related feedback effects due to dynamic hedging (Frey and Stremme, 1997; Platen and Schweizer, 1998; Schonbucher and Wilmott, 2000). Those models generally assume that hedging occurs instantaneously and increases stock price volatility. Because hedging is not instantaneous in actual trading, we assume that hedging occurs at a lag and so hedging affects the return autocorrelation.³

We test this autocorrelation prediction using data on options on individual stocks. We develop a measure of “hedging demand” that quantifies “if the underlying stock price increases by 1%, what fraction of the shares outstanding must option writers additionally purchase to remain hedged?”. Hedging demand reflects two components: The first component is the sensitivity of the hedge to changes in the underlying price. More specifically, this sensitivity is linked to the convexity of option payoff or “gamma,” in option terminology. In contrast, a purely linear financial derivative (e.g. a futures contract) has zero gamma and the hedge is constant. The second component is the number of options outstanding. Intuitively, hedging one hundred option contracts generates more price impact in the underlying stock than hedging one option contract.

In our empirical tests, we study the cross sectional variation in hedging demand. We show that stocks with higher hedging demand have higher return autocorrelations. As we move from the lowest to highest quintile of hedging demand, the daily return autocor-

²Many papers document that demand pressure affects the pricing of financial securities. For the stock market, see Shleifer (1986); Wurgler and Zhuravskaya (2002); Greenwood (2005). For the effect of demand pressure of options on their own prices, see Green and Figlewski (1999), Bollen and Whaley (2004), and Garleanu, Pedersen, and Poteshman (2009). Our paper studies the effect of options on stock prices.

³When it is costly to hedge continuously, Cetin et al (2006) note that traders often use Black-Scholes hedges at fixed intervals, as we model here. Traders can also improve their hedging with more nuanced methods (Leland, 1985; Boyle and Vorst, 1992; Cetin et al, 2006).

relation increases from -5.0% to -1.6%. The *marginal* effect of hedging demand on return autocorrelation is positive, but the *total* return autocorrelation is still negative because the effect of hedging demand is not enough to overcome the background negative autocorrelation in stock returns. Past research suggests that this background negative return autocorrelation is due to bid-ask bounce or compensation for liquidity provision.⁴ The effect of hedging demand on return autocorrelation is robust in different subsamples of our data (first half only, latter half only, large firms only, and excluding observations from the week of option expiration). In terms of the time horizon, the effect of hedging demand is statistically significant up to 100 trading days and the point estimate of its effect decays to zero at 325 trading days.

We also measure the economic magnitude using a portfolio sorting strategy. This portfolio sorting analysis complements the regressions, since portfolio sorting analyses are less sensitive than regressions to outliers and parametric misspecification. Since our theory suggests that stocks with higher hedging demand have higher autocorrelation, we sort stocks into quintiles based on the interaction of past returns and hedging demand. After controlling for the return factors RMRF, SMB, HML, UMD, and Short-Term Reversal, we find a gross alpha of 0.33% weekly (18% annualized) with a t-statistic of 4.2. While not the main focus of this paper, a back-of-the-envelope calculation suggests the portfolio sorting strategy survives transaction costs with a net alpha of 0.10% weekly (5.4% annualized).

So far, we have interpreted the results as hedging demand creating more return autocorrelation. However, a potential confound is that hedging demand may instead measure news entering the option market before the stock market. For example, informed traders might purchase options to make leveraged bets. At the same time, information could slowly integrate into the stock market, causing return autocorrelation (Hong and Stein, 1999). This potential confound is related to, but distinct from, research showing that various ratios related to option volume forecast future stock returns because news enters the option

⁴See Roll (1984); Lehmann (1990); Campbell, Grossman, and Wang (1993); Avramov, Chordia, Goyal (2006); Nagel (2012).

market first.⁵ One difference between those papers and our paper is that we focus on return autocorrelation, rather than just returns. Another difference is that those papers focus on option volume (i.e. flows) and the hedging theory presented here instead focuses on the total level of options that option writers must hedge.

We address this potential confound with an instrumental variable, the absolute difference between the underlying stock price and the nearest round number. This instrument for hedging demand is based on the institutional idiosyncrasy that exchange-traded options are struck at round numbers (e.g. \$600.00, rather than \$613.12). Furthermore, option convexity/gamma rises as the stock price approaches the option strike price. Therefore, the absolute difference between the underlying stock price and the nearest round number allows us to isolate variation in hedging demand, unrelated to news. In particular, our instrumental variable only moves the gamma of outstanding options, which is separate from variation in any option quantities, level or flow. The instrumental variable regression corroborates our other results that hedging demand creates more return autocorrelation.

The effect of hedging by option writers is related to the mechanical buying/selling by portfolio insurers, which the Presidential Task Force on Market Mechanisms (the “Brady Report”) argued exacerbated the October 1987 stock market crash (Brady, 1988).⁶ Portfolio insurers attempt to limit portfolio losses by selling as the stock price falls, which also generates an upward sloping demand curve for the stock. Typically, portfolio insurers implement their strategy using index futures, so as the price falls, portfolio insurers sell index futures and index arbitrageurs then transmit the effect to the individual stocks in the index.

Our analysis relies on the assumption that the end users in aggregate are option buyers and hence the option writers in aggregate provide liquidity by selling options and dynami-

⁵Examples include the put-call volume ratio (Easley, O’Hara, and Srinivas, 1998; Pan and Poteshman, 2006), the option-stock volume ratio (Johnson and So, 2012), and the delta-weighted option order imbalance (Hu, 2014). In contrast, Yang (2015) and Andreou et al (2014) argue that option quantities (volume and open-interest) measure belief disagreement and hence forecast negative returns due to the Miller (1977) effect.

⁶For overviews of other viewpoints on the 1987 crash, see the symposium in the Summer 1988 issue of the *Journal of Economic Perspectives*.

cally hedging their exposure. This assumption is closely related to the literature on demand effects in option pricing, which finds that hedged option writing on individual stocks earns positive returns (Bollen and Whaley, 2004; Frazzini and Pedersen, 2012; Cao and Han, 2013).⁷ For example, Frazzini and Pedersen (2012) argue that hedged option writing earns positive returns because end users value embedded leverage and hence are option buyers. Bakshi and Kapadia (2003) also find positive returns to hedged option writing on individual stocks, but instead attributes it to a fundamental negative volatility risk premium. Other papers on end user demand find that index options have a greater imbalance of end user demand than individual stock options, which helps explain differences in the pricing between the two types of options (Bollen and Whaley, 2004; Garleanu, Pedersen, and Poteshman, 2009).

The fact that hedged option writing earns positive returns may appear to be somewhat in tension with Lakonishok et al (2007), who find that nonmarket makers are net sellers of individual stock options. Their finding is largely driven by the net short position of clients of full-service brokerages, which they describe as including hedge fund clients of brokerages like Merrill Lynch. However, their proprietary dataset does not include data on whether or not the option traders are also hedging in the stock market. Hence, one cannot observe if the nonmarket makers (in particular, hedge funds) provide liquidity in the option market by opportunistically selling options and hedging their exposures. Because market makers have specific obligations to stand ready to buy and sell options at all times, market makers are not the sole liquidity providers in the option market.⁸ The mechanism we study, along with the papers on demand effects in option pricing, focuses on the hedging behavior of all liquidity providers, not just the formally designated market makers.

Our empirical work connects most directly to other empirical research on the effect of a derivative on its underlying asset. Figlewski and Webb (1993) argue that options

⁷Bollen and Whaley (2004) find that hedged option writing on individual stocks earns positive returns, but the standard errors are large, given their sample of 20 stocks. Using larger cross sections, Frazzini and Pedersen (2012) and Cao and Han (2013) find statistically significant profits. Frazzini and Pedersen (2012) and Cao and Han (2013) state their result as hedged option *buying* earns negative returns, so we state the equivalent result that hedged option *writing* earns positive returns.

⁸See CBOE (2009) for an example of the specific obligations of option market makers.

alleviate short-sale constraints and improve informational efficiency in the underlying stock. Other research focuses on effects at the introduction of a new derivative (Conrad, 1989; Detemple and Jorion, 1990; Sorescu, 2000). Ni, Pearson, and Poteshman (2005) and Golez and Jackwerth (2012) study re-hedging effects at option expiration. We study how the re-hedging of options exposure affects the underlying stocks, in general circumstances. Derivatives can also affect their underlying assets by changing the incentives of the asset owners (e.g. for empirical work on the “empty creditor problem” in credit markets, see Subrahmanyam, Tang, and Wang, 2014).⁹ While ETFs are not derivatives, hedging by liquidity providers of leveraged ETFs can similarly affect the underlying stock (Cheng and Madhavan, 2009). Ben-David, Franzoni, and Moussawi (2014) argue that un-leveraged ETFs increase volatility because they attract high-turnover investors.

At a thematic level, our paper also connects to papers on financial innovation. Recent theoretical work on financial innovation has emphasized the effect of derivatives on relaxing scarcity of the underlying (Banerjee and Graveline, 2014) and amplifying bets between agents with heterogeneous beliefs (Kubler and Schmedders, 2012; Simsek, 2013).

Section 1.2 describes the theoretical framework to motivate our measure of hedging demand. Section 1.3 describes our dataset and basic patterns in the time series and cross section. Section 1.4 describes the main empirical results (baseline results, robustness tests, long horizon effects, and decomposing returns into their systematic and idiosyncratic components). Section 1.5 measures the economic magnitude using a portfolio sorting strategy. Section 1.6 describes the instrumental variable regression. Section 1.7 concludes with a discussion of natural extensions of this paper.

⁹Like us, others also refer to the idea that derivatives can affect their underlying assets using the idiom of “the tail wagging the dog.” This group includes both researchers and policy makers (e.g. Finnerty and Park, 1987; Brown-Hruska, 2006; Subrahmanyam, Tang, and Wang, 2014).

1.2 Theoretical Framework and Hedging Demand

In our theoretical framework, we aggregate all the option writers into one representative agent option writer. In this section, we model how the option writer hedges, how the option writer re-hedges after changes in the underlying price, and how this re-hedging affects stock prices.

1.2.1 Notation and Option Terminology

Each stock has many different options on it, e.g. put/call, different maturities, and different strike prices. Let i denote the stock, t denote the time, and k denote the option. Then, let V_{itk} denote the value of option of type k (e.g. a call option with strike price \$600 that expires in 30 days). Let $written_{itk}$ denote the number of options of type k sold by the representative agent option writer. The option writer's aggregate exposure across the portfolio of options he has sold is then:

$$V_{i,t,agg} := \sum_k (written_{itk} \cdot V_{itk})$$

For example, if the option writer has sold one unit of the option $k = 1$ and one unit of the option $k = 2$, then $V_{i,t,agg} = V_{i,t,1} + V_{i,t,2}$.

We extensively use two concepts from the options literature: “delta” (Δ) and “gamma” (Γ). Delta and gamma characterize the sensitivity of an option or portfolio of options to the underlying price.

$$\begin{aligned}\Delta &:= \frac{\partial V}{\partial P} \\ \Gamma &:= \frac{\partial^2 V}{\partial P^2} = \frac{\partial \Delta}{\partial P}\end{aligned}$$

The sensitivity of the portfolio is a linear combination of the sensitivity of the individual options:

$$\begin{aligned}\Delta_{i,t,agg} &= \sum_k (written_{itk} \cdot \Delta_{itk}) \\ \Gamma_{i,t,agg} &= \sum_k (written_{itk} \cdot \Gamma_{itk})\end{aligned}$$

1.2.2 Hedging and Re-Hedging

Consider the representative agent option writer who has sold various options on a given stock i and wants to hedge his position. The option writer hedges his position because the end users in aggregate are option buyers. For simplicity, assume all options are European options. In the model, the option writer hedges his exposure on stock i with stock i itself. Hence, we drop the i subscript in the model for notational simplicity. In the empirical work, we focus on cross-sectional variation across stocks and hence we use the i subscript in the regressions.

The representative agent option writer W aims to collect the premium for writing options, as opposed to making any directional bet on the underlying stock. Let ξ_t^W denote the number of shares of the underlying stock that the option writer holds at time t . The option writer's objective function is to minimize the variance of his net worth N_{t+1}^W :

$$\min_{\xi_t^W} \text{Var}(N_{t+1}^W)$$

where

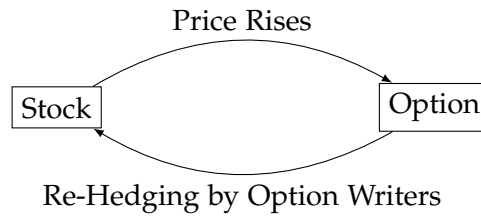
$$N_{t+1}^W = \underbrace{-V_{t+1,agg}}_{\text{Sold Options}} + \underbrace{\xi_t^W \cdot P_{t+1}}_{\text{Stock}} + \underbrace{(N_t^W + V_{t,agg} - \xi_t^W \cdot P_t)}_{\text{Cash}}$$

The option writer can only choose his exposure to the underlying stock ξ_t^W . The option writer cannot choose to dispose of the options and cannot choose the trivial solution of holding zero stock and zero options. Intuitively, while one option writer can sell his exposure to another option writer, the representative agent option writer hedges his exposure.

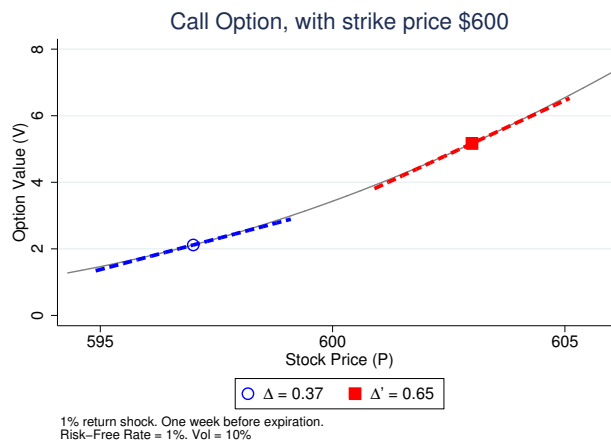
Proposition 1. *The solution to the option writer's objective function is to buy $\xi_t^W = \Delta_{t,agg}$ units of the underlying.*

Such a position is often called a “delta-neutral position,” because hedge offsets the delta of the options sold. Intuitively, the option writer's wealth is now unaffected by small changes in the price of the underlying (Figure 1.1b).

(a) Diagram Depicting Basic Mechanism



(b) Upward Sloping Demand Curve



(c) Price Impact of Re-Hedging Causes Return Autocorrelation

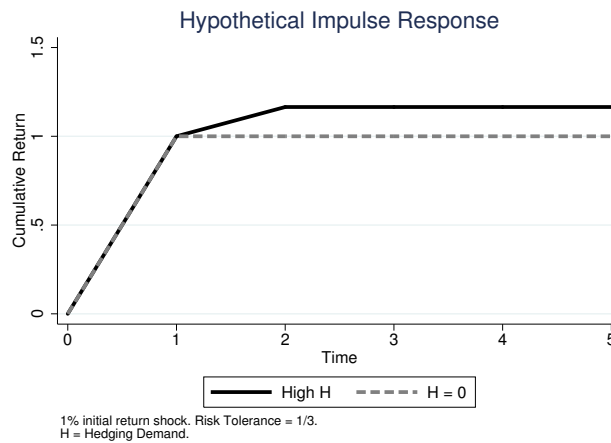


Figure 1.1: *How Hedged Option Writing Affects Stock Prices*

These figures overview the theory we describe in Section 1.2. Dynamic hedging by option writers creates an upward sloping demand curve for the underlying stock and return autocorrelation.

$$\begin{aligned}
\frac{\partial N^W}{\partial P} &= \text{Loss from Sold Options} + \text{Gain from Hedge} \\
&= -\frac{\partial V_{agg}}{\partial P} + \Delta_{agg} = 0
\end{aligned}$$

Proposition 2. *If the underlying price rises by 1%, the option writer must buy an additional $\Gamma_{agg} \cdot 1\%P$ shares of the underlying to remain hedged.*

If the underlying price P rises by 1%, then the option writer now demands approximately $\Delta_{agg} + \frac{\partial \Delta_{agg}}{\partial P} \cdot 1\%P$ shares of the underlying stock (Figure 1.1b). Since $\Gamma_{agg} = \frac{\partial \Delta_{agg}}{\partial P}$, the *change in the hedge* is $\Gamma_{agg} \cdot 1\%P$. If $\Gamma_{agg} > 0$, then this is an upward sloping demand curve. As the price rises, hedging requires buying more of the underlying.

An option writer hedging a put option has the same upward sloping demand curve. For put option, we have $\Delta_{put} < 0$, so the option writer hedges his position by “buying Δ_{put} ” (i.e. “selling $|\Delta_{put}|$ ”) units of the underlying stock. Since put options are also convex ($\Gamma_{put} > 0$), as the underlying price rises, the option writer buys more of the underlying stock (i.e. reduces his short position).

Above, we assumed that the option writer hedges by dynamically trading the underlying stock. One could alternately model the option writer selling many options on the different firms in an index and then attempting to hedge his exposure with the index itself. However, this hedge is imperfect because options have non-linear payoffs: an option on a portfolio is distinct from a portfolio of options. As the underlying stock prices change, the required amount of stock in each company (to remain hedged) can deviate from the weights in the index. For example, suppose the index is just an equal-weighted average of two stocks: X and Y . Assume X and Y have the same number of shares outstanding, so we only need to track the share prices. Suppose the option writer sells one call option on each stock. If X rises by \$10 and Y falls by \$10, then the index is unchanged. However, the option writer now needs more of stock X and less of stock Y . Hence, the weights in his hedge now deviate from the weights in the index. For this reason, we model hedging with the stock itself.

1.2.3 Price Impact of Re-Hedging

Next, we describe how re-hedging by the option writer increases autocorrelation in the stock returns. There are two assets in this model: the risky asset (“stock”) and the risk-free asset. At some final date T , the stock pays a single terminal dividend of $F_T = F_0 + \sum_{j=0}^T \epsilon_j$, where $\epsilon_j \sim N(0, 1)$. We normalize the stock to have a total quantity of 1. We also normalize the net risk-free rate to 0, by assuming the risk-free asset is elastically supplied at that rate.

There are three groups of representative agents: the option buyer, the option writer, and the “fundamental investor.” In the option market, the option buyer purchases options from the option writer. For simplicity, assume that the option buyer does not interact with the stock market.¹⁰ The option writer sells the options and hedges his risk in the stock market.

In the stock market, the representative agent option writer interacts with the representative agent fundamental investor (FI). The fundamental investor’s role is simply to provide a downward sloping demand curve, against which the option writer trades. The following setup is one way to generate this downward-sloping demand curve: The fundamental investor only trades based on one-period ahead fundamentals F_{t+1} . This assumption prevents the fundamental investor from front running the option writer’s hedging. The fundamental investor has constant absolute risk aversion (CARA) utility, where τ denotes the risk tolerance, i.e. reciprocal of risk aversion. Let N_{t+1}^{FI} denote her net worth in the next period and ζ_t^{FI} denote the fundamental investor’s demand for the underlying stock. Given CARA utility and normally distributed risk, her objective function is equivalent to mean-variance optimization:

$$\max_{\zeta_t^{FI}} E[N_{t+1}^{FI}] - \frac{1}{2\tau} \text{Var}[N_{t+1}^{FI}]$$

where $N_{t+1}^{FI} = F_{t+1}\zeta_t^{FI} + (N_t^{FI} - P_t\zeta_t^{FI})$. This setup implies that the fundamental investor has a downward sloping demand curve for the underlying stock $\zeta_t^{FI} = \tau \cdot (E[F_{t+1}] - P_t)$.

As a benchmark, consider the equilibrium with only the fundamental investor, i.e. no

¹⁰One can alternately allow the option buyer to trade in the stock market, as long as he does not fully offset the hedging from the option writer.

hedging by the option writer. In this benchmark, the stock price is simply the fundamental value less a risk premium, $P_t^{bmark} = F_0 + \sum_{j=1}^t \epsilon_j - \frac{1}{\tau}$. Therefore, when there is only the fundamental investor, stock prices are a random walk:

$$\begin{aligned} Cov(P_{t+1}^{bmark} - P_t^{bmark}, P_t^{bmark} - P_{t-1}^{bmark}) &= Cov(\epsilon_{t+1}, \epsilon_t) \\ &= 0 \end{aligned}$$

Now, we add the hedging by the option writer. We assume that the option writer can only hedge every other period $t \in \{0, 2, 4, \dots\}$, e.g. because it is too costly to hedge every period. Therefore, when the option writer delta hedges, his demand for the underlying stock is $\tilde{\zeta}_t^W = \Delta_{t,agg}$ if $t \in \{0, 2, 4, \dots\}$ or $\tilde{\zeta}_t^W = \Delta_{t-1,agg}$ if $t \in \{1, 3, 5, \dots\}$. Setting supply equal to demand, we get the following prices:

$$P_t = \begin{cases} P_t^{bmark} + \frac{\Delta_{t,agg}}{\tau} & \text{if } t \in \{0, 2, 4, \dots\} \\ P_t^{bmark} + \frac{\Delta_{t-1,agg}}{\tau} & \text{if } t \in \{1, 3, 5, \dots\} \end{cases}$$

When the option writer hedges (in the even periods $t \in \{0, 2, 4, \dots\}$), we must solve for a fixed point because $\Delta_{t,agg}$ is a function of P_t . This fixed point exists as long as $\Gamma_{agg}/\tau < 1$. Intuitively, as the option writer buys more of the underlying stock, the price rises, which induces more desire for shares, and so on, until we reach the fixed point solution.¹¹ The condition $\Gamma_{agg}/\tau < 1$ ensures this process converges. We see that if the fundamental investor has infinite risk tolerance τ , then prices are equal to the random-walk benchmark. Intuitively, this is because higher risk tolerance lowers the price impact of the hedging by the option writer.

To illustrate this equilibrium, consider the impulse response where there is a single positive news shock at time 1 (i.e. $\epsilon_j = 0$ for $j \neq 1$). Figure 1.1c depicts this impulse response. Using the fact that $\Delta_2 \approx \Delta_0 + \Gamma_1(P_2 - P_0)$, we can solve for the closed form

¹¹ Alternately, instead of assuming that the option writer can only hedge every other period, one could assume that the option writer hedges every period, but does so in a backward-looking manner. That is, one could assume that the option writer holds $\Delta_{t-1,agg}$ shares of the underlying at period t . This alternate setup avoids the fixed point and predicts more sluggish price adjustment. However, this alternate setup has the drawback that the option writer is no longer forward-looking.

Table 1.1: *Closed Form Solution of Impulse Response*

Returns	With Option Writers	Benchmark
$P_1 - P_0$	ϵ_1	ϵ_1
$P_2 - P_1$	$\epsilon_1 \cdot \frac{\Gamma_{1,agg}/\tau}{1 - \Gamma_{1,agg}/\tau}$	0
$P_{t+1} - P_t$ for $t \geq 2$	0	0

solution displayed in Table 1.1.

Proposition 3. *Hedging by the option writer creates return autocorrelation, $Cov(P_2 - P_1, P_1 - P_0) > 0$. As $\Gamma_{1,agg}$ rises, return autocorrelation rises. Also, as the fundamental investor's risk tolerance τ falls, return autocorrelation rises.*

Since we normalized $Var(\epsilon_1) = 1$, we have $Cov(P_2 - P_1, P_1 - P_0) = \frac{\Gamma_{1,agg}/\tau}{1 - \Gamma_{1,agg}/\tau} > 0$. The key ratio is $\Gamma_{1,agg}/\tau$, which is the ratio of the option writer's hedging demand to the fundamental investor's risk tolerance. Intuitively, as $\Gamma_{1,agg}$ rises, autocorrelation rises because the option writer must buy more shares to re-hedge. And, as the fundamental investor's risk tolerance τ falls, autocorrelation rises because there is more price impact from the hedging.

The model also predicts that the delta hedging affects the stock price permanently, as opposed to causing a temporary "overshooting." For example, in the impulse response, the returns after period 2 (i.e. $P_{t+1} - P_t$ for $t \geq 2$) are zero, not negative. Intuitively, at period 2, the option writer increases his holdings of the underlying stock because his option exposure has increased. However, after period 2, because there are no more shocks, the option writer remains hedged and does not need to adjust his stock holdings.

1.2.4 Hedging Demand and its Empirical Counterpart

We now define our key variable "hedging demand." Our model showed that higher Γ_{agg} is associated with higher serial correlation, holding all other variables fixed. Intuitively, when Γ_{agg} is higher, for the same shock, the option writer must buy more of the underlying stock.

Hedging demand is Γ_{agg} re-scaled so we can compare across different stocks with different share prices and shares outstanding.

Re-scaling is also necessary because our price impact model used a setup with CARA utility and normally distributed risk. This setup is convenient for models with heterogeneous agents because the demand for risky assets is independent of wealth. Hence, the equilibrium does not depend on the distribution of wealth amongst the agents. However, one drawback of this setup is that the product of two normal random variables is not normally distributed. Therefore, in these types of models, “returns” are typically just the change in the price level $P_{t+1} - P_t$. However, in empirical work, returns are $\frac{P_{t+1} - P_t}{P_t}$, so we re-scale to account for this difference.

Definition 1. *Theoretical Version of Hedging Demand: If the price of the underlying stock increases by 1%, what fraction of the shares outstanding must the option writer additionally purchase to remain hedged? Answer:*

$$\begin{aligned} H^* &:= \frac{1\%P}{\text{SharesOut}} \cdot \Gamma_{agg} \\ &= \frac{1\%P}{\text{SharesOut}} \cdot \sum_k (\text{written}_k \cdot \Gamma_k) \end{aligned}$$

Definition 2. *Empirical Counterpart of Hedging Demand:*

$$\begin{aligned} H_{it} &:= \frac{1\%P_{it}}{\text{SharesOut}_{it}} \cdot \sum_k (100 \cdot OI_{itk} \cdot \Gamma_{itk}) \\ \text{Log}H_{it} &:= \text{Log}(H_{it}) \end{aligned}$$

where i denotes the stock, t denotes the time, and k denotes the option.

We denote the theoretical version H^* with an asterisk because the empirical counterpart of hedging demand contains two adjustments. First, in actual trading, each option contract corresponds to 100 shares. Hence, we include 100 as a normalizing constant. Second, we approximate written_{itk} with open interest OI_{itk} . In our model, the two are equivalent (i.e. $\text{written}_{itk} = OI_{itk}$) because the representative agent option writer hedges all the sold options.

However, in a more nuanced model, there may be an additional group of option

writers who are end users and do not hedge their exposure. Let $written'_{itk}$ denote the options sold by these option writers who do not hedge. (Let $written_{itk}$ continue to denote the options sold by the option writers who do hedge.) In this more nuanced model, $OI_{itk} = written_{itk} + written'_{itk}$. If a constant fraction of the open interest is always actively hedged, then our empirical definition $LogH_{it}$ equals the “true” log hedging demand plus a constant. For this reason, we focus on $LogH_{it}$ in our empirical work.

The other reason we focus on $LogH_{it}$ is because the log form allows us to more easily control for the model of price pressure. Price impact may depend on some combination of fraction of shares outstanding, fraction of volume, etc. By using the log form, log shares outstanding enters linearly into our estimation. We then add controls of log market cap, log dollar volume, and log share price. It also turns out that using just hedging demand H_{it} gives similar empirical results.

1.3 Dataset and Summary Statistics

We use panel data on two levels: the option level and the equity level. The option-level data aggregates to equity-level data. At both levels, data frequency is daily. Our dataset spans from Jan 1996 to Aug 2013 for a total of roughly 4400 days. Table 1.2 contains a list of the main variables used in this paper.

At the option level, an observation is an equity-date-option triplet. For example, an observation might be the Apple, Inc. call option with strike price \$600.00 that expires on Feb 16, 2013 as observed on Jan 02, 2013. Our option-level data come from OptionMetrics. We include all observations from the start of the database to the latest available data, Jan 1996 to Aug 2013. In terms of cross-sectional span, the OptionMetrics database covers all U.S. exchange-listed options. While OptionMetrics also has data for index options, in this paper, we focus on options on individual stocks. Our option-level data is 118 GB in size and contains 977 million observations.

For option risk sensitivities gamma Γ , we use the estimates from OptionMetrics, which are based on the Cox, Ross, and Rubinstein (1979) binomial tree method. Their model

accommodates discrete and continuous dividends. The other key option-level datum is open interest, which is the total number of contracts that are not settled. Our main explanatory variable Hedging Demand (H_{it}) collapses these option-level data into an equity-level statistic at the daily frequency.

At the equity level, an observation is an equity-date pair. For example, an observation might be Apple, Inc. on Jan 02, 2013. Across the 17 years, our dataset of 977 million option-level observations rolls up into 4.3 million equity-level observations, when merged with equity-level data. The mean number of equities per daily cross section is roughly 1000. Using CRSP, we obtain equity-level data on holding period returns, shares outstanding, trading volume, close prices, and bid-ask spreads. From CRSP, we also obtain lower-frequency data on CRSP size decile cutoffs. From Compustat, we obtain data on accounting book value. We lag lower-frequency data to avoid look-forward bias.

In Section 1.5, we analyze our results using a long-short portfolio sorting strategy. In that section, we further collapse the data to the day-level. For example, an observation might be the return spread for the high vs low quintile of stocks on Jan 02, 2013. For that analysis, we also use data from Ken French's Data Library for the factors of RMRF, HML, SMB, UMD, and Short-Term Reversal.

1.3.1 Time Series and Cross Sectional Patterns in Our Dataset

Figure 1.2 plots the time series and cross sectional patterns of hedging demand. Figure 1.2a plots the average of log hedging demand $\text{Log}H$ over time. The large decline in hedging demand in 2008-2009 is due to both a decline in the open interest of options and a large decline in stock prices. A decline in open interest lowers hedging demand because it reduces the number of options the option writers must hedge. A large decline in stock prices lowers hedging demand because it reduces the gamma, the sensitivity of the hedge to changes in the underlying stock price. In particular, options are generally issued with strike prices close to the current stock price. Hence, when there is a large price rise or fall, the stock price is now far away from the strike price of previously issued options (Figure 1.4). As gamma is

Table 1.2: *List of Key Variables*

This table lists the key variables used in this paper.

Variable	Symbol	Description
Hedging Demand	H	Proxy for “If the underlying stock price rises by 1%, what fraction of the shares outstanding must the option writer additionally purchase to remain hedged?” See Section 1.2.4 for formal definition.
Lagged Moving Average of Variable Y	Y_{ma} (i.e. subscript ma)	Lagged moving average over the last five trading days (one trading week).
Option Delta (subscript k for individual option and agg for portfolio)	Δ	First derivative of an option (or portfolio of options) with respect to underlying price, i.e. $\frac{\partial V}{\partial S}$. For individual options, we use the estimates computed by Option Metrics, which is based on the Cox, Ross, and Rubinstein (1979) binomial tree model.
Option Gamma (subscript k for individual option and agg for portfolio)	Γ	Second derivative of an option (or portfolio of options) with respect to underlying price, i.e. $\frac{\partial^2 V}{\partial S^2}$. For individual options, we use the estimates computed by Option Metrics, which is based on the Cox, Ross, and Rubinstein (1979) binomial tree model.
Open Interest	OI	Total number of options contracts that are currently not settled (“open”).
Simple Return	R	Net simple returns, including dividends.
Log Return	r	$r = \text{Log}(1 + R)$.
Excess Log Return	r^e	Log return in excess of the log risk-free rate, $r_{it}^e = r_{it} - r_{ft}$. For the risk-free rate, we use the one-month Treasury bill.
Market Cap	$MktCap$	Total equity market capitalization of a firm, using close prices.
Dollar Volume	$DVolume$	Total dollar volume for a given stock on a given day.
Share Price	P	Unadjusted price per share on the close of each day.
CAPM Beta	β	We estimate betas using the Scholes and Williams (1977) method to account for potentially nonsynchronous trading. We compute the betas quarterly and use the estimates from the previous quarter to avoid look-forward bias.
Book/Market Ratio	$Book/Mkt$	(Book Equity)/(Market Capitalization). We lag by two quarters to ensure that the accounting data was already publicly available on each date. Following Fama and French (1993), we define Book Equity = Stockholder’s Equity + Deferred Taxes and Investment Tax Credit (if available) - Preferred Stock.
Bid-Ask Spread	$Spread$	(Ask - Bid)/(Midpoint)
Turnover	$Turnover$	(Share Volume)/(Shares Outstanding)

highest near the strike price (i.e. near the “kink” in the option payoff diagram), large price movements reduce the gamma of previously issued options and reduce hedging demand.

Hedging demand has a strong monthly cyclic variation due to options expiration. Figure 1.2a also shows average log hedging demand for Jan 2012 to Aug 2013 (end of our dataset). Exchange traded equity options expire the Saturday after the third Friday of the expiration month. For example, all exchange-traded equity options that expired in January 2012, expired on Saturday Jan 21, 2012. We observe that open interest falls dramatically right before each month’s expiration. This monthly periodicity offers an alternate interpretation to the results of Ni, Pearson, and Poteshman (2005), which finds that stock prices cluster near the strike price on option expiration dates. In theory, delta hedging by option buyers should cause clustering and delta hedging by option writers should cause de-clustering. However, when Ni, Pearson, and Poteshman (2005) roughly classify traders into delta hedgers and end users, they find clustering for both groups. As a result, they conclude that effects other than re-hedging are at work. The alternate explanation suggested by Figure 1.2a is that the option writers are the delta hedgers and they have the fewest number of option contracts to hedge right before option expiration (due to the monthly periodicity). Therefore, in the time series, there would be the less de-clustering (i.e. more clustering) at expiration. This prediction is a time series prediction, which is separate from the cross sectional variation we use in the main results of this paper.

Figure 1.2b plots the cross sectional dispersion of $\text{Log}H$. Specifically, it displays the cross sectional standard deviation over time and displays a histogram of de-meanned $\text{Log}H$. In the histogram, we de-mean $\text{Log}H$ because average $\text{Log}H$ changes over time. We observe that the standard deviation of $\text{Log}H$ is higher in more recent years.

Table 1.3 displays the summary statistics. Panel (a) displays the summary statistics for the entire dataset. Panel (b) shows how the averages vary across the quintiles of hedging demand. Roughly speaking, our dataset is 1000 firms per day across 4400 days. The firms in our sample are relatively large, with a mean market capitalization of \$7.9 billion and a median market capitalization of \$2.2 billion. The variable $H \cdot \frac{\text{ShareOut}}{\text{Volume}}$ re-scales hedging

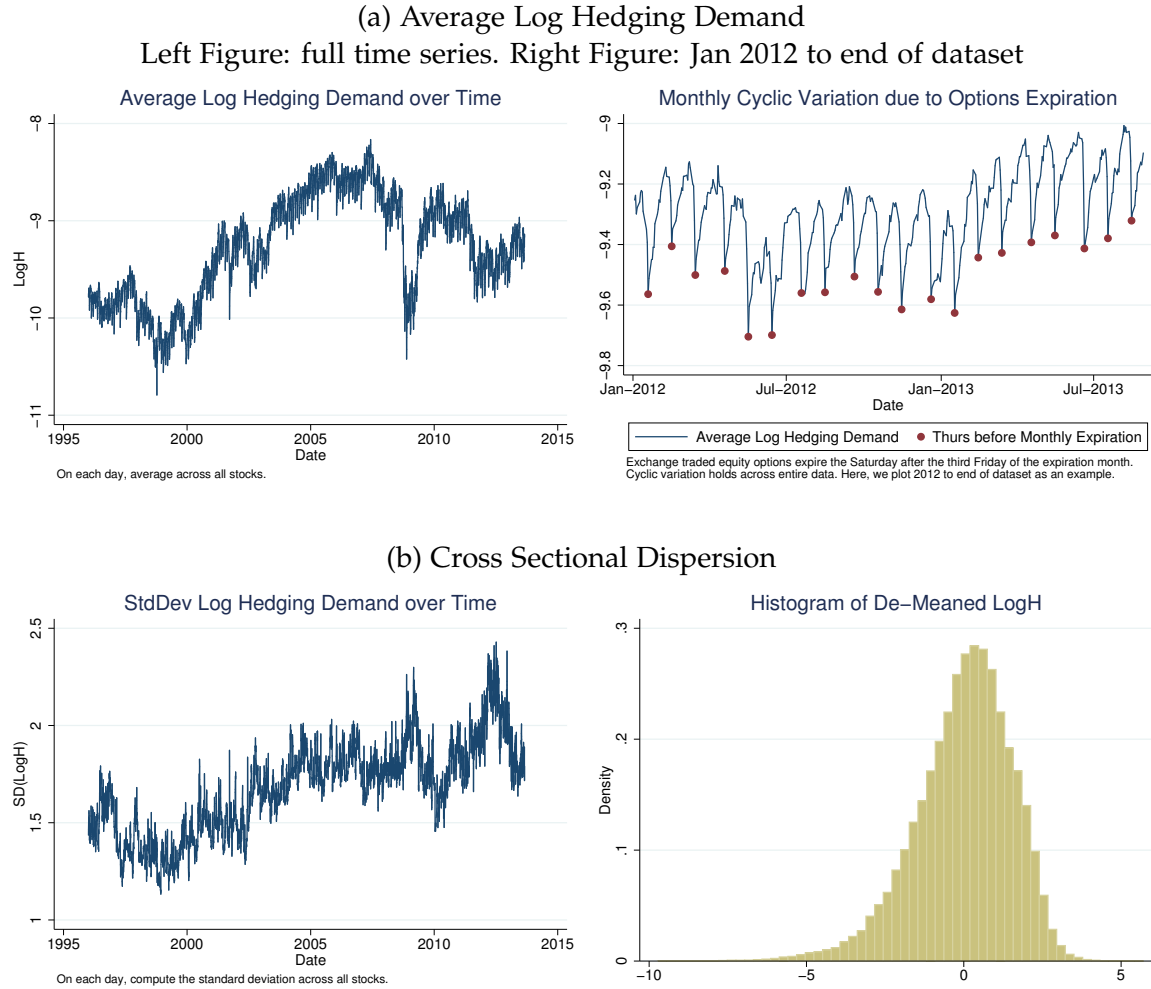


Figure 1.2: *Time Series and Cross Sectional Properties of Log Hedging Demand*

We define hedging demand as $H_{it} := \frac{1\%P_{it}}{\text{SharesOut}_{it}} \cdot \sum_k (100 \cdot OI_{itk} \cdot \Gamma_{itk})$ where P_{it} is the stock price, OI_{itk} is the open interest of option k , and Γ_{itk} is the convexity/gamma of option k . These graphs plot the log of hedging demand ($\text{Log}H_{it}$), which is the key explanatory variable in our analysis. While the time series patterns in Panel (a) convey a general sense of log hedging demand, our analysis relies on the cross sectional variation in log hedging demand in Panel (b). See also Table 1.3 for summary statistics by quintile of hedging demand.

demand by daily volume instead of shares outstanding. Hence, it measures the fraction of *daily volume* the option writer would need purchase to remain hedged after a 1% increase in the underlying stock price. Because we use open interest as a proxy for the number of options being hedged, both H and $H \cdot \frac{\text{ShareOut}}{\text{Volume}}$ are an upper bound on the hedging demand—specifically, the hedging demand if all open option contracts were actively hedged by option writers.

Average hedging demand H is 0.028% and average $H \cdot \frac{\text{ShareOut}}{\text{Volume}}$ is 4.24%, which means a 1% increase in the underlying stock price would require purchasing an additional 0.028% of the shares outstanding (4.24% of the daily volume). For the highest quintile, $H = 0.09\%$ and $H \cdot \frac{\text{ShareOut}}{\text{Volume}} = 10.08\%$. Stocks with higher hedging demand are larger and more liquid, i.e. higher turnover and lower spread. CAPM beta rises with hedging demand. The mean of volatility does not rise with hedging demand, but the median (not shown) rises slightly.

1.4 Main Empirical Results

The theory in Section 1.2 predicts that stocks with higher hedging demand have higher return autocorrelation, due to hedging by option writers. In this section, we test this prediction using a regression framework. In addition to the baseline regressions, we also analyze the robustness to different subsamples and different estimation methods; the effect on long horizon returns; and the effect of decomposing returns into their systematic and idiosyncratic components.

1.4.1 Baseline Regression

Let r_{it}^e denote the log returns of equity i at time t in excess of the log risk-free rate. For the risk-free rate, we use the one-month Treasury bill. The subscript ma denotes a lagged moving average over the last five trading days (i.e. one trading week). For example, $r_{ma,i,t}^e$ denotes the lagged moving average of returns. In our baseline regression, the right hand side regressors are returns $r_{ma,i,t}^e$, log hedging demand $\text{Log}H_{ma,i,t}$, controls $X_{ma,i,t}$, and their

Table 1.3: Summary Statistics

This table displays summary statistics of our main variables. We express hedging demand H in percent for legibility. We include $LogH_{ma}$ and de-meanned $LogH_{ma}$, since we use those values in the text. The average of the de-meanned $LogH_{ma}$ is not exactly 0 because we de-mean cross sectionally, as opposed to over the entire dataset. The variable $H \cdot \frac{ShareOut}{Volume}$ is hedging demand scaled by volume, instead of shares outstanding. We abbreviate market capitalization as “MktCap” and dollar volume as “DVolume.” See Table 1.2 for full list of variable definitions. $N = 4,326,618$.

(a) Full Dataset					
	mean	sd	p10	p50	p90
Year	2004.8	5.24	1997	2005	2012
Num Firms per Day	987.0	124.3	828	1027	1105
H , in %	0.028	0.044	0.0012	0.011	0.073
$H \cdot \frac{ShareOut}{Volume}$, in %	4.20	5.72	0.28	2.13	10.5
$LogH_{ma}$	-9.22	1.64	-11.3	-9.11	-7.24
$LogH_{ma}$, de-meanned	-0.0063	1.56	-2.00	0.15	1.82
MktCap, in billions	7.94	21.0	0.40	2.25	17.6
DVolume, in millions	49.3	122.5	1.40	14.0	119.1
Turnover, in %	0.84	0.90	0.16	0.55	1.81
Spread, in %	0.65	1.11	0.026	0.15	1.87
β	1.08	8.84	0.23	1.00	2.07
Book/Market Ratio	0.60	0.55	0.18	0.49	1.08
Volatility, in %	2.37	1.56	1.03	1.97	4.10

(b) By Quintile of Hedging Demand					
	1	2	3	4	5
H , in %	0.00	0.01	0.01	0.03	0.09
$H \cdot \frac{ShareOut}{Volume}$, in %	0.80	1.75	3.11	5.38	10.05
$LogH_{ma}$	-11.53	-9.91	-9.05	-8.31	-7.27
$LogH_{ma}$, de-meanned	-2.31	-0.69	0.16	0.91	1.94
MktCap, in billions	2.78	4.01	6.27	13.47	13.27
DVolume, in millions	9.28	18.73	33.21	72.74	113.69
Turnover, in %	0.52	0.65	0.78	0.91	1.34
Spread, in %	0.81	0.69	0.63	0.58	0.56
β	1.00	1.06	1.05	1.12	1.18
Book/Market Ratio	0.71	0.63	0.58	0.53	0.53
Volatility, in %	2.37	2.31	2.32	2.35	2.51

mean coefficients

interactions.

$$\begin{aligned}
r_{i,t+1}^e = & b_{0,t} + b_1 \cdot r_{ma,i,t}^e + \lambda \cdot r_{ma,i,t}^e \times \text{Log}H_{ma,i,t} \\
& + b_2 \cdot \text{Log}H_{ma,i,t} + \underbrace{b_3 \cdot r_{ma,i,t}^e \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t}}_{\text{Controls}} + \epsilon_{i,t+1}
\end{aligned} \tag{1.1}$$

To focus on cross sectional variation, we allow for a different intercept $b_{0,t}$ for each time period. Our main regressions use the Fama-MacBeth methodology, which implicitly includes the time varying intercept; our panel regressions explicitly include the time fixed effect.

Return autocorrelation is the ability to forecast future returns using past returns. In Equation 1.1, the regressors involving past returns are $r_{ma,i,t}^e$, $r_{ma,i,t}^e \times \text{Log}H_{ma,i,t}$, and $r_{ma,i,t}^e \times X_{ma,i,t}$. Therefore, the *total* return autocorrelation is $b_1 + \lambda \cdot \text{Log}H_{ma,i,t} + b_3 \cdot X_{ma,i,t}$. We cross-sectionally de-mean log hedging demand $\text{Log}H_{ma,i,t}$ and the controls $X_{ma,i,t}$, so the coefficient b_1 has the natural interpretation as the return autocorrelation for a firm with average characteristics (i.e. mean log hedging demand, mean log market capitalization, etc.).

The coefficient of interest is the λ coefficient on the interaction term $r_{ma,i,t}^e \times \text{Log}H_{ma,i,t}$. This coefficient measures the *marginal* effect of hedging demand on return autocorrelation. The theory in Section 1.2 predicts that $\lambda > 0$, i.e. higher $\text{Log}H_{ma,i,t}$ is associated with more autocorrelation.

Since we estimate the impact of hedging demand on autocorrelation, the key controls $X_{ma,i,t}$ of interest are other variables that might also affect autocorrelation or price impact (e.g. size of the firm). We control for these covariates by adding them as interactions with $r_{ma,i,t}^e$. Strictly speaking, since we control for multiple covariates, we should express $X_{ma,i,t}$ as a matrix. However, we keep it in this form for simplicity.

The lagged moving average allows us to estimate the average effect over multiple days because there is no fundamental reason to believe that hedging must happen within one day. Market practitioners, with whom we spoke, confirmed that they often rebalance over a few days to reduce transaction costs. In contrast, Cheng and Madhavan (2009) study leveraged ETFs, which must contractually rebalance at the end of the day.

In a more generalized regression, one could alternately estimate a separate coefficient

for each lag. However, we use the lagged moving averages because it is easier to display the results concisely. Note also that we use the average of the log, i.e. $\text{Log}H_{ma,i,t}$, as opposed to the log of the average, i.e. $\text{Log}(H_{ma,i,t})$. We use the former because it more easily generalizes to explicitly estimating a separate coefficient for each j th lag of $\text{Log}(H_{i,t-j})$. Both the average of the log and the log of the average give very similar estimates.

Table 1.4 displays the baseline results, estimated using the Fama-MacBeth methodology. The controls are log market capitalization, log dollar volume, and log share price. Since $\text{Log}H = \text{Log}(\text{Num Shares to Remain Hedged}) - \text{Log}(\text{SharesOut})$, these controls are also equivalent to normalizing hedging demand by some linear combination of log market capitalization, log dollar volume, and log share price. Since log turnover equals log dollar volume minus log market capitalization, if we alternately use log turnover in our controls, we find that autocorrelation falls as turnover increases, which is consistent with the literature (e.g. Campbell, Grossman, and Wang, 1993). Later, in Table 1.8, we use alternate estimation methods, including Fama-MacBeth with Newey-West standard errors and panel regression with clustered standard errors. For legibility, displayed coefficients are regression coefficients multiplied by 100.

Table 1.4 Column (1) shows the regression of returns on lagged returns. Columns (2), (3), (4), and (5) then estimate the effect of hedging demand (the λ coefficient on the interaction term $r_{ma,i,t}^e \times \text{Log}H_{ma,i,t}$) with different controls. We see that the b_1 coefficient on r_{ma}^e is negative throughout. This negative coefficient matches the general finding that equity returns have negative autocorrelation in the cross section at short horizons, which past research has linked to bid-ask bounce or compensation for liquidity provision (e.g. Roll, 1984; Nagel, 2012).

Next, we examine the λ coefficient on the interaction term $r_{ma,i,t}^e \times \text{Log}H_{ma,i,t}$ across Columns (2), (3), (4), and (5). In terms of statistical significance, the estimates are all significant with t-statistics exceeding 3.0. The estimated λ coefficient is positive, which implies that higher hedging demand is associated with higher return autocorrelation. This finding matches the prediction from the theory in Section 1.2. The estimated economic

Table 1.4: Baseline Results

This table displays the baseline results. We are interested in the coefficient λ on the interaction term $r_{ma,i,t}^e \times \text{Log}H_{ma,i,t}$, where r^e denotes excess log returns and $\text{Log}H$ denotes log hedging demand. This coefficient λ estimates the additional return autocorrelation associated with an increase in hedging demand. Subscript ma (e.g. $\text{Log}H_{ma}$) denotes a lagged moving average over the last five trading days. We cross-sectionally de-mean $\text{Log}H_{ma,i,t}$ and the controls $X_{ma,i,t}$, so the b_1 coefficient measures the return autocorrelation for average firm. Data are 1996 to 2013, daily. See Table 1.2 for full list of variable definitions.

$$r_{i,t+1}^e = b_{0,t} + b_1 \cdot r_{ma,i,t}^e + \lambda \cdot r_{ma,i,t}^e \times \text{Log}H_{ma,i,t} + b_2 \cdot \text{Log}H_{ma,i,t} + b_3 \cdot r_{ma,i,t}^e \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t} + \epsilon_{i,t+1}$$

Regression methodology is Fama-MacBeth. For legibility, displayed coefficients are regressions estimates multiplied by 100. For brevity, we do not display the non-interacted controls (“main effects”).

VARIABLES	(1)	(2)	(3)	(4)	(5)
r_{ma}^e	-1.39*** (-3.1)	-1.47*** (-3.3)	-2.32*** (-5.1)	-2.68*** (-5.8)	-3.15*** (-7.0)
$r_{ma}^e \times \text{Log}H_{ma}$		0.61*** (3.5)	0.94*** (5.5)	0.72*** (3.8)	0.79*** (4.1)
$\text{Log}H_{ma}$		-0.01*** (-3.4)	-0.01*** (-3.5)	-0.01*** (-5.0)	-0.01*** (-5.3)
$r_{ma}^e \times \text{LogMktCap}_{ma}$			-1.10*** (-5.9)	-1.64*** (-4.6)	-1.90*** (-5.2)
$r_{ma}^e \times \text{LogDVolume}_{ma}$				0.63* (1.9)	0.61* (1.8)
$r_{ma}^e \times \text{LogPrice}_{ma}$					0.59 (1.4)
Observations	4,326,618	4,326,618	4,326,618	4,326,618	4,326,618
R-squared	0.017	0.028	0.043	0.056	0.068
Number of groups	4,434	4,434	4,434	4,434	4,434
Main effects	Y	Y	Y	Y	Y

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

magnitude depends on the controls, but is roughly similar across the columns.

In Table 1.4 Column (5), which includes all the controls of log market capitalization, log dollar volume, and log share price, the marginal impact of log hedging demand is $\hat{\lambda} = 0.79\%$. The standard deviation of the de-meaned $LogH_{ma}$ is 1.5, so a one standard deviation increase in log hedging demand increases the return autocorrelation by 1.1%. This increase is meaningful as the average firm has an estimated return autocorrelation of $\hat{b}_1 = -3.15\%$. Put differently, as we move from the lowest to highest quintile of hedging demand, the total return autocorrelation ($b_1 + \lambda \cdot LogH_{ma,i,t}$) increases from -5.0% to -1.6%. The total autocorrelation is still negative for the highest quintile because our effect is not enough to offset the fact that stocks have reversion in general at this time scale ($b_1 < 0$). However, in other asset classes with less background reversion, hedging option writing could push the total autocorrelation positive.

While not the focus of this paper, our tables show that $LogH_{ma}$ robustly predicts lower returns in the cross section, i.e. $b_2 < 0$. This effect is driven by the fact that option open interest is a strong negative predictor of cross sectional returns (Yang, 2014). This finding is related to the research showing that high option *volume* forecasts low returns (Johnson and So, 2012).

1.4.2 Robustness: Different Subsamples and Estimation Methods

In this subsection, we explore the robustness of the baseline results to different subsamples and different estimation methods. To fix a point of reference, we compare relative to the main baseline regression in Table 1.4 Column (5). Naturally, one can choose a different point of reference as well. In our main baseline regression, we found $\hat{\lambda}_{baseline} = 0.79\%$ with t-statistic of 4.1. For ease of reference, we repeat the results of that regression estimation in each robustness table. As before, we are interested in the λ coefficient on the interaction term $r_{ma,i,t}^e \times LogH_{ma,i,t}$. This coefficient λ estimates the additional return autocorrelation associated with an increase in log hedging demand. Also, as before, for legibility, displayed coefficients are regressions estimates multiplied by 100.

Table 1.5: Effect on Daily Turnover

This table displays the relationship between hedging demand and daily turnover. The variable $|r^e|$ denotes absolute value of the excess log returns and $LogH$ denotes log hedging demand. Subscript ma (e.g. $LogH_{ma}$) denotes a lagged moving average over the last five trading days. We cross-sectionally de-mean $LogH_{ma,i,t}$ and the controls $X_{ma,i,t}$. Data are 1996 to 2013, daily. See Table 1.2 for full list of variable definitions.

$$\begin{aligned} turnover_{i,t+1} = & b_{0,t} + b_1 \cdot |r^e_{ma,i,t}| + \textcolor{red}{c} \cdot |r^e_{ma,i,t}| \times LogH_{ma,i,t} \\ & + b_2 \cdot LogH_{ma,i,t} + b_3 \cdot |r^e_{ma,i,t}| \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t} + \epsilon_{i,t+1} \end{aligned}$$

For legibility, displayed coefficients are regressions estimates multiplied by 100. Controls: log market capitalization, log dollar volume, log share price, and their interactions with $|r^e|$.

VARIABLES	(1) Turnover	(2) Turnover
$ r^e_{ma} $	5.07*** (13.5)	5.07*** (7.6)
$ r^e_{ma} \times LogH_{ma}$	2.86*** (11.3)	2.86*** (8.8)
$LogH_{ma}$	0.02*** (10.4)	0.02*** (6.6)
Observations	4,326,618	4,326,618
R-squared	0.447	0.447
Controls + xterms	Y	Y
StdErr Cluster	Time	Time + Firm
Fixed Effect	Time	Time

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 1.6 illustrates that the main baseline result is robust in various subsamples of our main dataset. Column (1) repeats the main baseline regression from Table 1.3. The rest of the columns display the different subsample regressions. In Column (2) and Column (3), we verify that the effect exists in both the first half and second half of the sample. One could potentially be concerned that option behavior has shifted over time. For example, dynamic hedging techniques have improved over time, as participants lower the price impact of hedging. Furthermore, as discussed by Campbell et al (2001), idiosyncratic volatility has risen over time, which affects the pricing of options on individual stocks. Column (2) uses the first half of the sample, i.e. before 2005. The effect is somewhat weaker in the first half. However, it is still similar in magnitude $\hat{\lambda}^{pre2005} = 0.72\%$ and is still significant with a t-statistic of 2.3. Column (3) uses the second half of the data, i.e. 2005 and later. The economic magnitude increases to $\hat{\lambda}^{post2005} = 0.86\%$ and t-statistic of 3.9.

Column (4) uses the subsample of firms with market capitalization in the top 50% of the CRSP universe, i.e. larger firms. In our 1996-2013 sample, that corresponds to \$11.8 billion (average market capitalization) and \$1.4 billion (average minimum required market capitalization). The economic magnitude is larger among these firms with $\hat{\lambda}^{bigfirms} = 0.85\%$ (t-statistic = 3.6).

Table 1.8 illustrates that the main baseline result is robust to different estimation methods. In this table, the estimation method varies across columns. Columns (1), (2), and (3) compare variants of the Fama-MacBeth estimation method. Column (1) repeats the main baseline result from Table 1.3. Column (2) shows a Fama-MacBeth regression with additional controls of beta, log book-to-market ratio, and log spread. These additional controls have only a small effect on economic magnitude and statistical significance of the main baseline result.

Column (3) shows the main baseline regression using the Fama-MacBeth methodology, but with Newey-West standard errors with lag length of 20 trading days (i.e. one trading month). Newey-West standard errors adjust for time series correlation in the error terms, using a triangle (Bartlett) kernel for the correlation structure. The key parameter in the Newey-West procedure is the lag length. In this dataset, lag lengths of 1 trading day to 120

Table 1.6: *Baseline Results, Using Different Subsamples of the Main Dataset*

This table estimates the baseline results using different subsamples of the main dataset. We are interested in the coefficient λ on the interaction term $r_{ma,i,t}^e \times \text{Log}H_{ma,i,t}$, where r^e denotes excess log returns and $\text{Log}H$ denotes log hedging demand. Column (1) is the baseline regression from Table 1.4 Column (5), which uses the full dataset. Column (2) uses the first half of our dataset, i.e. before 2005. Column (3) uses the second half of our dataset, i.e. 2005 and later. Column (4) uses the subsample of firms with market capitalization in the top 50% of the CRSP universe on each day, i.e. larger firms. In our dataset, these larger firms have a mean market capitalization of \$11.8 billion. Data are 1996 to 2013, daily. See Table 1.2 for full list of variable definitions.

$$r_{i,t+1}^e = b_{0,t} + b_1 \cdot r_{ma,i,t}^e + \lambda \cdot r_{ma,i,t}^e \times \text{Log}H_{ma,i,t} + b_2 \cdot \text{Log}H_{ma,i,t} + b_3 \cdot r_{ma,i,t}^e \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t} + \epsilon_{i,t+1}$$

Regression methodology is Fama-MacBeth. For legibility, displayed coefficients are regressions estimates multiplied by 100. Controls: log market capitalization, log dollar volume, log share price, and their interactions with returns.

VARIABLES	(1)	(2)	(3)	(4)
r_{ma}^e	-3.15*** (-7.0)	-3.83*** (-6.4)	-2.45*** (-3.7)	-3.46*** (-6.8)
$r_{ma}^e \times \text{Log}H_{ma}$	0.79*** (4.1)	0.72** (2.3)	0.86*** (3.9)	0.85*** (3.6)
$\text{Log}H_{ma}$	-0.01*** (-5.3)	-0.01*** (-4.7)	-0.01*** (-2.8)	-0.00 (-0.9)
Observations	4,326,618	1,993,710	2,332,908	2,784,652
R-squared	0.068	0.068	0.069	0.076
Number of groups	4,434	2,256	2,178	4,434
Controls + xterms	Y	Y	Y	Y
Dataset	All	Pre-2005	Post-2005	Size>p50

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 1.7: *Baseline Results, Using Different Subsamples of Options*

This table estimates the baseline results using different subsamples of the options. We are interested in the coefficient λ on the interaction term $r_{ma,i,t}^e \times \text{Log}H_{ma,i,t}$, where r^e denotes excess log returns and $\text{Log}H$ denotes log hedging demand. Column (1) is the baseline regression from Table 1.4 Column (5), which uses the full dataset. Column (2) omits days during the week of option expiration. Column (3) only uses data from put options. Column (4) only uses data from call options. Data are 1996 to 2013, daily. See Table 1.2 for full list of variable definitions.

$$r_{i,t+1}^e = b_{0,t} + b_1 \cdot r_{ma,i,t}^e + \lambda \cdot r_{ma,i,t}^e \times \text{Log}H_{ma,i,t} + b_2 \cdot \text{Log}H_{ma,i,t} + b_3 \cdot r_{ma,i,t}^e \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t} + \epsilon_{i,t+1}$$

Regression methodology is Fama-MacBeth. For legibility, displayed coefficients are regressions estimates multiplied by 100. Controls: log market capitalization, log dollar volume, log share price, and their interactions with returns.

VARIABLES	(1)	(2)	(3)	(4)
r_{ma}^e	-3.15*** (-7.0)	-3.89*** (-7.6)	-3.18*** (-7.1)	-3.23*** (-7.2)
$r_{ma}^e \times \text{Log}H_{ma}$	0.79*** (4.1)	0.77*** (3.5)		
$r_{ma}^e \times \text{Log}H_{ma}^{put}$			1.39*** (8.5)	
$r_{ma}^e \times \text{Log}H_{ma}^{call}$				0.36* (1.9)
$\text{Log}H_{ma}$	-0.01*** (-5.3)	-0.01*** (-4.2)		
$\text{Log}H_{ma}^{put}$			-0.01*** (-6.2)	
$\text{Log}H_{ma}^{call}$				-0.01*** (-4.6)
Observations	4,326,618	3,311,810	4,266,788	4,314,979
R-squared	0.068	0.068	0.068	0.068
Number of groups	4,434	3,393	4,434	4,434
Controls + xterms	Y	Y	Y	Y
Dataset	All	Omit Expir Wk	Only Puts	Only Calls

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 1.8: *Baseline Results, Using Different Estimation Methods*

This table displays the effect of varying the estimation method. The key coefficient of interest is λ . Column (1) is the baseline regression from Table 1.4 Column (5), which uses the Fama-Macbeth procedure ("FM" in the table). Column (2) is a Fama-MacBeth regression with the additional controls of beta, log book-to-market ratio, and log spread. Column (3) the baseline regression using Fama-MacBeth with Newey-West standard errors, lag length of 20 trading days. Column (4) is a panel regression with fixed effects by time and standard errors clustered by time. Column (5) clusters standard errors by time and firm. The r-squared of the panel regressions does not include the time fixed effect and hence one cannot directly compare it to the r-squared of the Fama-MacBeth regressions. Data are 1996 to 2013, daily. See Table 1.2 for full list of variable definitions.

$$r_{i,t+1}^e = b_{0,t} + b_1 \cdot r_{ma,i,t}^e + \lambda \cdot r_{ma,i,t}^e \times \text{Log}H_{ma,i,t} + b_2 \cdot \text{Log}H_{ma,i,t} + b_3 \cdot r_{ma,i,t}^e \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t} + \epsilon_{i,t+1}$$

For legibility, displayed coefficients are regressions estimates multiplied by 100. Controls: log market capitalization, log dollar volume, log share price, and their interactions with returns.

VARIABLES	(1)	(2)	(3)	(4)	(5)
r_{ma}^e	-3.15*** (-7.0)	-3.69*** (-9.1)	-3.15*** (-7.0)	-3.96*** (-4.4)	-3.96*** (-4.1)
$r_{ma}^e \times \text{Log}H_{ma}$	0.79*** (4.1)	0.68*** (3.4)	0.79*** (4.7)	1.35*** (5.5)	1.35*** (5.0)
$\text{Log}H_{ma}$	-0.01*** (-5.3)	-0.01*** (-4.9)	-0.01*** (-5.2)	-0.02*** (-7.5)	-0.02*** (-7.4)
Observations	4,326,618	4,326,618	4,326,618	4,326,618	4,326,618
R-squared	0.068	0.112	0.068	0.001	0.001
Number of groups	4,434	4,434	4,434	4,434	4,434
Controls + xterms	Y	Y + Extras	Y	Y	Y
StdErr Cluster	FM	FM	FM+NW	Time	Time+Firm
Fixed Effect	FM	FM	FM	Time	Time

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

trading days give similar t-statistics, so we simply choose the round number of 20 trading days (i.e. one trading month). In Column (3), we see that the t-statistic actually increases from 4.1 to 4.7 after applying Newey-West standard errors. This is because the errors are slightly negatively correlated up to two trading months, so accounting for the negative autocorrelation improves the t-statistic. In any case, the effect is modest since our dependent variable is (non-overlapping) returns and the usual concern with returns is cross-sectional correlation, not time series correlation.

Column (4) shows a panel regression with fixed effects by time and standard errors clustered by time. This panel regression is similar to the Fama-MacBeth estimation procedure. The main difference is that the panel regression more heavily weights cross sections with more observations and cross sections where there is more heterogeneity in the explanatory variables. In our dataset, this translates to more heavily weighting the more recent observations, as there are more observations per cross section in recent years and more dispersion in $\text{Log}H$ in recent years (Figure 1.2). Since hedging demand has a stronger effect in the second half of the sample, this econometric logic predicts the panel estimates should exceed the Fama-MacBeth estimates, which is what we find.

The panel regression results are significantly stronger in both economic magnitude and statistical significance. The economic magnitude of the panel estimate is $\hat{\lambda}^{panel} = 1.35\%$, roughly 1.7x the Fama-MacBeth estimates. The t-statistic of the panel estimate is 5.5, which is significantly larger than the t-statistic for the Fama-MacBeth estimates.

Column (5) shows a panel regression with fixed effects by time and standard errors clustered by *time and firm*. Clustering by firm is similar to Newey-West standard errors. The main distinction is that the Newey-West procedure assumes that the serial correlation decays over time, whereas clustering by firm does not. As discussed in Petersen (2009), if we set the Newey-West lag length to $T - 1$, the formula for the standard error is the same as clustering by firm, modulo a weighting function. Additionally clustering by firm only affects the t-statistics modestly. As before, this is because the general concern for (non-overlapping) returns is usually cross-sectional correlation, not time series correlation.

1.4.3 Effect on Long Horizon Returns

To analyze how hedging affects return autocorrelation at long horizons, we replace the dependent variable in Equation 1.1 with $\sum_{j=1}^N r_{i,t+j}^e$, the N period cumulative return.

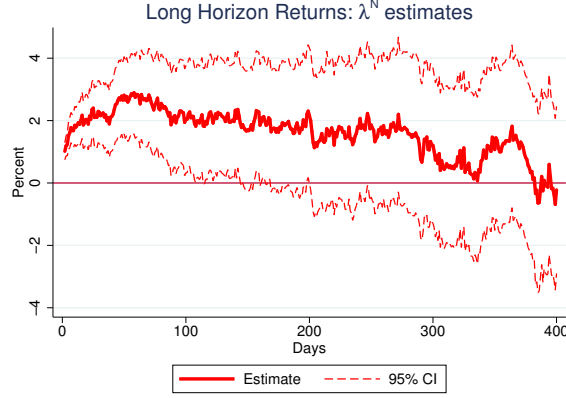
$$\begin{aligned} \sum_{j=1}^N r_{i,t+j}^e = & b_{0,t} + b_1 \cdot r_{ma,i,t}^e + \lambda^N \cdot r_{ma,i,t}^e \times \text{Log}H_{ma,i,t} \\ & + b_2 \cdot \text{Log}H_{ma,i,t} + b_3 \cdot r_{ma,i,t}^e \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t} + \epsilon_{i,t+N} \end{aligned} \quad (1.2)$$

The coefficient of interest is λ^N , the effect of the interaction of past returns with hedging demand on the future N period cumulative return. Since the dependent variable is now overlapping returns, we must account for serial correlation in the error term. We use Newey-West standard errors with a lag length of 400 trading days, since that is the maximum horizon we consider. For reference, our total dataset has roughly 4400 trading days.

Figure 1.3 displays two graphs. Figure 1.3a plots the λ^N estimate at different horizons N . The effect of hedging on long horizon returns is statistically significant up to 100 trading days and the point estimate is positive up to 325 trading days. Figure 1.3b expresses the same information in the form of an impulse response. It plots the total effect of a positive 1% return shock on the highest and lowest quintile of hedging demand. Because of stocks in general have reversion at this time scale ($b_1 < 0$), the total autocorrelation remains negative for the highest quintile. However, we clearly see that the highest quintile has less reversion than the lowest quintile.

This persistence is in line with the theory. In Section 1.2, we showed that after a positive shock to the stock price, option writers choose to hold more stock to re-hedge and continue to hold that amount of stock as long as there are no additional price shocks. The effect after option expiration depends on the option buyers' actions. If the option buyer holds the stock after expiration, then the effect will persist. (In our sample, average days to maturity, weighted by open interest, is around 100 trading days.) Our data are for exchange-traded equity options, which are "physically delivered" not "cash settled" (CBOE, 2014). Hence, at expiration, the option writer delivers the underlying shares to the option buyer and the

(a) λ^N Estimates at Different Horizons



(b) Impulse Response: Highest vs Lowest Quintile of Hedging Demand

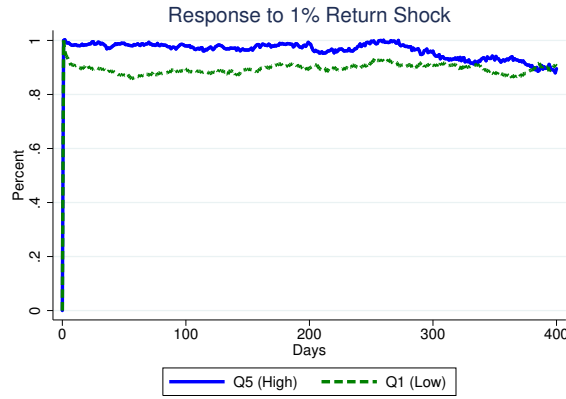


Figure 1.3: Impact at Long Horizons

These graphs illustrate the relationship between log hedging demand and long horizon returns. The dependent variable is the long horizon return $\sum_{j=1}^N r_{i,t+j}^e$. We are interested in the coefficient λ^N on the interaction term $r_{ma,i,t}^e \times \text{Log}H_{ma,i,t}$, where r^e denotes excess log returns and $\text{Log}H$ denotes log hedging demand. We estimate the regression:

$$\begin{aligned} \sum_{j=1}^N r_{i,t+j}^e = & b_{0,t} + b_1 \cdot r_{ma,i,t}^e + \lambda^N \cdot r_{ma,i,t}^e \times \text{Log}H_{ma,i,t} \\ & + b_2 \cdot \text{Log}H_{ma,i,t} + b_3 \cdot r_{ma,i,t}^e \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t} + \epsilon_{i,t+N} \end{aligned}$$

Newey-West standard errors with a lag length of 400 trading days. A panel regression with clustering by time and firm and time fixed effects gives similar results.

persistence we observe suggests that the option buyers continue to hold these delivered shares for some time. The effect eventually decays at long horizons, which suggests that (1) the option buyers eventually sell the stock or (2) arbitrage capital slowly enters the stock market, which increases the risk-tolerance τ of the fundamental investors (Duffie, 2010).

1.4.4 Systematic vs Idiosyncratic Components of Returns

We examine whether the effect of hedging demand comes from the systematic component of returns or the residual idiosyncratic component of returns. We decompose returns using the CAPM and compute betas using the Scholes and Williams (1977) method. The Scholes and Williams (1977) method accounts for potentially non-synchronous trading, as that can create biased estimates due to measurement error. We compute the betas quarterly and use the estimates from the previous quarter to avoid look-forward bias. We can then decompose the returns into the systematic component and the residual idiosyncratic component:

$$\begin{aligned} r_i^{e,sys} &= \beta \cdot r_{mkt}^e \\ r_i^{e,idio} &= r_i^e - r_i^{e,sys} \end{aligned}$$

Hence, adding in the lagged moving averages, our regression is:

$$\begin{aligned} r_{i,t+1}^e &= b_{0,t} + b_1^{sys} \cdot r_{ma,i,t}^{e,sys} + b_1^{idio} \cdot r_{ma,i,t}^e \\ &\quad + \lambda^{sys} \cdot r_{ma,i,t}^{e,sys} \times \text{Log}H_{ma,i,t} + \lambda^{idio} \cdot r_{ma,i,t}^{e,idio} \times \text{Log}H_{ma,i,t} \\ &\quad + b_2 \cdot \text{Log}H_{ma,i,t} + b_3 \cdot r_{ma,i,t}^e \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t} + \epsilon_{i,t+1} \end{aligned} \quad (1.3)$$

We are interested in the estimates for the coefficients λ^{sys} and λ^{idio} on the interaction terms $r_{ma,i,t}^{e,sys} \times \text{Log}H_{ma,i,t}$ and $r_{ma,i,t}^{e,idio} \times \text{Log}H_{ma,i,t}$. These coefficients estimate the additional return autocorrelation associated with an increase in log hedging demand, decomposed into the systematic and idiosyncratic component of returns.

Table 1.9 displays the results of our decomposition regression. For reference, Column (1) repeats the main baseline result from Table 1.4, where we found that $\hat{\lambda} = 0.79\%$. Column (2) shows the decomposed results. We see that $\hat{\lambda}^{sys}$ is statistically insignificant, with a t-statistic

of 1.1. On the other hand, $\hat{\lambda}^{idio} = 0.90\%$ with a t-statistic of 4.7. The standard errors on the systematic component are large enough that we cannot reject the hypothesis $\hat{\lambda}^{sys} = \hat{\lambda}^{idio}$.

From this decomposition, we can conclude that the idiosyncratic component of returns drives at least part of the effect of hedging demand. As for the systematic component of returns, there are two possibilities: The first is that the systematic component has no effect. The second is that the systematic component has a partial effect, but we simply do not have enough statistical power to detect it as $StdError(\hat{\lambda}^{sys})$ is large in our regression. This standard error is large because our regressions rely on cross-sectional variation. Most of the cross-sectional variation comes from the idiosyncratic component and so the effect of the systematic component is poorly estimated.

1.5 Portfolio Sorting Strategy

As an alternate way to quantify the economic magnitude of hedging demand by option writers, we analyze the returns to a portfolio sorting strategy. This methodology expresses the economic magnitude in terms of an annualized alpha, which may be easier to grasp intuitively given the large literature using the portfolio sorts. We sort stocks into quintiles based on $r_{ma,i,t}^e \times LogH_{ma,i,t}$ and analyze the portfolio returns of each quintile. Controlling for the factors RMRF, SMB, HML, UMD, and Short-Term Reversal, we find that stocks in the highest quintile outperform stocks in the lowest quintile by 0.33% alpha per week (t-statistic of 4.1). On an annualized basis, that is equivalent to gross alpha of 18%

The portfolio sorting methodology also helps address potential concerns with the Fama-MacBeth regressions of Section 1.4. First, the Fama-MacBeth methodology is sensitive to outliers. In contrast, the portfolio sorting methodology dampens the effect of outliers by sorting stocks into quintiles. Second, the Fama-MacBeth methodology is sensitive to regression misspecification. The portfolio sorting strategy (in particular, the per quintile regressions) allow us to check for potential non-monotonic relationships in the alphas. Another difference is that in the portfolio sorting methodology, one controls for “covariances,” instead of “characteristics” (Daniel and Titman, 1997).

Table 1.9: Decomposing Returns into Systematic and Idiosyncratic Components

This table displays the effect of decomposing returns into a systematic and an idiosyncratic component. Column (1) shows the main baseline result from Table 1.4 Column (5) for reference. Column (2) shows the decomposed results. Using CAPM β , we decompose r_i^e into the systematic component ($r_i^{e,sys} = \beta \cdot r_{mkt}^e$) and the residual idiosyncratic component ($r_i^{e,idio} = r_i^e - r_i^{e,sys}$); in the regressions, we use lagged moving averages of this decomposition. We are interested in the coefficients λ^{sys} and λ^{idio} on the interaction terms $r_{ma,i,t}^{e,sys} \times \text{Log}H_{ma,i,t}$ and $r_{ma,i,t}^{e,idio} \times \text{Log}H_{ma,i,t}$. These coefficients estimate the additional return autocorrelation associated with an increase in log hedging demand, decomposed into the systematic and the idiosyncratic component of returns. Data are 1996 to 2013, daily. See Table 1.2 for full list of variable definitions.

$$\begin{aligned}
 r_{i,t+1}^e = & b_{0,t} + b_1^{sys} \cdot r_{ma,i,t}^{e,sys} + b_1^{idio} \cdot r_{ma,i,t}^{e,idio} \\
 & + \lambda^{sys} \cdot r_{ma,i,t}^{e,sys} \times \text{Log}H_{ma,i,t} + \lambda^{idio} \cdot r_{ma,i,t}^{e,idio} \times \text{Log}H_{ma,i,t} \\
 & + b_2 \cdot \text{Log}H_{ma,i,t} + b_3 \cdot r_{ma,i,t}^e \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t} + \epsilon_{i,t+1}
 \end{aligned}$$

Regression methodology is Fama-MacBeth. For legibility, displayed coefficients are regressions estimates multiplied by 100. Controls: log market capitalization, log dollar volume, log share price, and their interactions with returns.

VARIABLES	(1)	(2)
r_{ma}^e	-3.15*** (-7.0)	
$r_{ma}^{e,sys}$		-0.09 (-0.0)
$r_{ma}^{e,idio}$		-3.55*** (-8.4)
$r_{ma}^e \times \text{Log}H_{ma}$	0.79*** (4.1)	
$r_{ma}^{e,sys} \times \text{Log}H_{ma}$		14.45 (1.1)
$r_{ma}^{e,idio} \times \text{Log}H_{ma}$		0.90*** (4.7)
$\text{Log}H_{ma}$	-0.01*** (-5.3)	-0.01*** (-2.6)
Observations	4,326,618	4,326,618
R-squared	0.068	0.087
Number of groups	4,434	4,434
Controls + xterms	Y	Y

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Our Fama-MacBeth regressions used the interaction term $r_{ma,i,t}^e \times \text{Log}H_{ma,i,t}$ to measure the additional return autocorrelation associated with log hedging demand. Hence, we use that interaction term to sort stocks into quintiles. We skip one day between portfolio formation and the portfolio holding period, to avoid potential concerns about different closing times across markets. Our (non-overlapping) portfolio holding period is weekly.¹² That is, each week, we sort stocks into quintiles using the value of $r_{ma,i,t}^e \times \text{Log}H_{ma,i,t}$ at the close on Thursday. When there are market holidays, we adjust correspondingly. We then measure the returns over the following week.

For quintile i , let $R_{Qi,t+1}^e = R_{Qi,t+1} - R_{f,t+1}$ denote the excess returns for that portfolio. We use simple returns, as that is the standard in this literature. Portfolio returns are value-weighted. We test if the return spread between the highest and lowest quintiles persists after controlling for the standard return factors f_{t+1} , using the regression:

$$R_{Q5,t+1} - R_{Q1,t+1} = \alpha + \beta' \cdot f_{t+1} + \epsilon_{t+1} \quad (1.4)$$

Table 1.10 shows the results of this regression under various factor controls. The constant term estimates the alpha of the portfolio sort strategy. All the alphas are statistically significant with t-statistics exceed 4.0. Harvey, Liu, and Zhu (2014) argue that data mining is a concern in empirical asset pricing and recommend that 3.0 should be the new minimum required t-statistics for tests of expected returns in the cross section. Our portfolios here pass that test.

The return spread between the high vs low quintile survives the addition of the various factors. Column (1) does not include any factor controls and hence estimates there is a 0.52% difference in weekly returns (i.e. 30% annualized). Column (2) controls for the market factor (RMRF). Column (3) controls for the Fama-French 3-factor model (RMRF, SMB, HML). Column (4) controls for the Carhart 4-factor model (RMRF, SMB, HML, UMD).

In Column (5), we further add the Short-Term Reversal Factor. Our sorting variable is $r_{ma,i,t}^e \times \text{Log}H_{ma,i,t}$, which is an interaction term involving lagged returns. Hence, it is

¹²Using daily holding periods gives slightly stronger results.

Table 1.10: *Alpha of a Long-Short Portfolio (Weekly Holding Periods)*

This table displays the returns to a portfolio sorting strategy, where we sort stocks into quintiles based on the interaction term $r_{ma,i,t}^e \times \text{Log}H_{ma,i,t}$. The constant term estimates the alpha of the portfolio sorting strategy. We skip one day between portfolio formation and the portfolio holding period, to avoid potential concerns about different closing times across markets. Portfolios are held for one week, so the dependent variable is the (non-overlapping) weekly return spread between the highest quintile and the lowest quintile, $R_{Q5,t+1} - R_{Q1,t+1}$. Portfolios are value-weighted. Explanatory variables are the vector of the standard factors, f_{t+1} . Column (1) is the raw average return spread, i.e. without controlling for any return factors. Column (2) controls for the market factor. Column (3) controls for the Fama-French 3-factor model. Column (4) controls for the Carhart 4-factor model. Column (5) adds the Short-Term Reversal Factor. Data frequency is weekly so each 0.01% weekly alpha translates into roughly 0.50% annual alpha, before transactions costs.

$$R_{Q5,t+1} - R_{Q1,t+1} = \alpha + \beta' \cdot f_{t+1} + \epsilon_{t+1}$$

Newey-West standard errors with lag length of four trading weeks (one trading month).

VARIABLES	(1) Q5-Q1	(2) Q5-Q1	(3) Q5-Q1	(4) Q5-Q1	(5) Q5-Q1
RMRF, in %		0.11* (1.8)	0.11* (1.8)	0.14** (2.1)	-0.01 (-0.1)
SMB, in %			-0.00 (-0.0)	-0.02 (-0.2)	-0.04 (-0.5)
HML, in %			-0.00 (-0.0)	0.05 (0.4)	0.14 (1.3)
UMD, in %				0.10* (1.7)	0.16** (2.5)
ST Reversal, in %					0.50*** (7.9)
Constant	0.52*** (6.7)	0.50*** (6.5)	0.50*** (6.4)	0.48*** (6.1)	0.33*** (4.1)
Observations	916	916	916	916	916
R-squared	0.000	0.014	0.014	0.022	0.179

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

important to verify that the alpha is robust to controlling for the Short-Term Reversal Factor. By comparison, in the Fama-MacBeth baseline regressions in Section 1.4, we controlled for r_{ma}^e directly in the regression. After adding the Short-Term Reversal Factor, we find that the weekly alpha is 0.33% with t-statistic of 4.1. In economic magnitude, that is equivalent to 18% annualized alpha, before transactions costs.

While not the main focus of this paper, a back-of-the-envelope calculation suggests our portfolio sorting strategy survives transactions costs. On average, about 75% of the stocks change quintiles each week. The median bid-ask spread is 0.15%. (The mean bid-ask spread is higher, but in our sample that is driven by times such as the fall of 2008.) Therefore, one might estimate the round-trip transactions costs as $0.225\% = 0.15\% \times 2 \times 75\%$. Our portfolio sorting strategy therefore has a net weekly alpha of $0.105\% = 0.33\% - 0.225\%$, which is a net alpha of 5.4% annualized. This calculation does not include the cost of price impact, but more complex trading methods could further reduce the transactions cost.

In cross-sectional asset pricing, it is also standard to verify that the alphas are monotonic across the quintiles. Unless one has strong theoretical reasons for non-monotonicity, finding non-monotonic alphas suggests a potentially spurious relationship. Table 1.11 breaks down the returns by quintile. Specifically, we run the regression:

$$R_{Qi,t+1}^e = \alpha + \beta' \cdot f_{t+1} + \epsilon_{t+1}$$

We see that the alphas are indeed monotonic across the quintiles, as desired.

1.6 Identification via Instrumental Variable

The results from Section 1.4 and Section 1.5 show that higher log hedging demand is associated with higher return autocorrelation. However, to establish causality, we require exogenous variation. One potential confound is that the earlier results measure news entering option markets first, instead of the effect of hedging. When open interest increases, hedging demand also increases because there are more options to hedge. However, open interest also increases when informed traders purchase options to make leveraged bets.

Table 1.11: *Alpha per Quintile (Weekly Holding Periods)*

This table examines the returns to each separate quintile of the portfolio sorting strategy. In contrast to Table 1.10, here, the dependent variable is the (non-overlapping) excess weekly return of quintile i , $R_{Qi,t+1}^e$. We wish to verify that the alphas are monotonic across the quintiles. The constant term estimates the alpha of each portfolio. We sort equities into quintiles based on the interaction term $r_{ma,i,t}^e \times \text{Log}H_{ma,i,t}$. We skip one day between portfolio formation and the portfolio holding period, to avoid potential concerns about different closing times across markets. Portfolios are held for one week and are value-weighted. Explanatory variables are the vector of the standard factors, f_{t+1} . Column (1) - Column (5) show the results for each quintile's excess returns. Data frequency is weekly so each 0.01% weekly alpha translates into roughly 0.50% annual alpha, before transactions costs.

$$R_{Qi,t+1}^e = \alpha + \beta' \cdot f_{t+1} + \epsilon_{t+1}$$

Newey-West standard errors with lag length of four trading weeks (one trading month).

VARIABLES	(1) Q1	(2) Q2	(3) Q3	(4) Q4	(5) Q5
RMRF, in %	1.05*** (39.1)	0.98*** (60.0)	0.94*** (50.5)	0.96*** (40.3)	1.05*** (26.7)
SMB, in %	-0.02 (-0.3)	-0.12*** (-3.9)	-0.16*** (-4.4)	-0.12*** (-2.7)	-0.06 (-1.0)
HML, in %	0.13** (2.5)	0.24*** (7.0)	0.21*** (5.7)	0.27*** (5.4)	0.28*** (3.7)
UMD, in %	-0.16*** (-4.6)	-0.04* (-1.7)	-0.02 (-0.6)	0.07** (2.4)	-0.00 (-0.1)
ST Reversal, in %	-0.21*** (-5.6)	-0.11*** (-3.7)	0.00 (0.1)	0.09** (2.2)	0.28*** (6.6)
Constant	-0.02 (-0.4)	0.04 (1.1)	0.08*** (2.6)	0.13*** (3.8)	0.31*** (5.3)
Observations	916	916	916	916	916
R-squared	0.802	0.880	0.887	0.846	0.801

t-statistics in parentheses

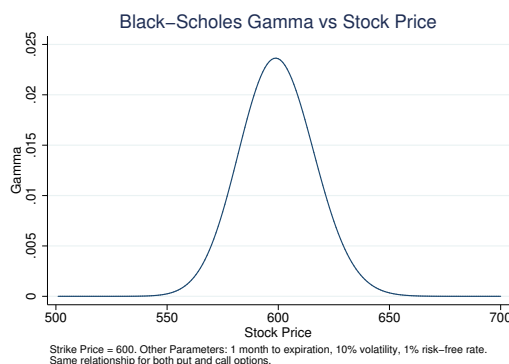
*** p<0.01, ** p<0.05, * p<0.1

If the information slowly integrates into the stock market, then we will observe return autocorrelation (Hong and Stein, 1999). As discussed in Section 1.1, this potential confound is related to, but distinct from, research showing that various ratios related to option volume forecast stock returns because news enters the option market first (Easley, O'Hara, and Srinivas, 1998; Pan and Poteshman, 2006; Johnson and So, 2012; Hu, 2014).

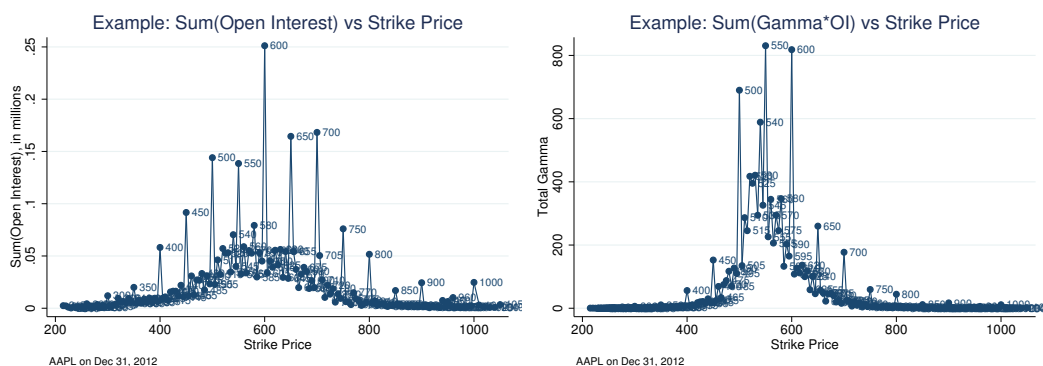
To address this potential confound, we use the absolute distance of the underlying stock price to nearest round number as our instrumental variable for hedging demand. Our instrument focuses on variation in the convexity/gamma of outstanding options. In contrast, the news-entering-option-markets-first confound focuses on changes in the open interest. The instrument relies on two facts about options: (1) For a put or call option, gamma is highest when the underlying stock price equals the strike price on the option contract. For example, Figure 1.4a depicts this relationship for an option with strike price \$600 that expires in one month. Intuitively, the strike price is the “kink” of the option payoff and so that is where the gamma is the highest. In the limit, on expiration day, gamma is infinite at the strike price. (2) As a matter of institutional detail, option exchanges only issue options with strike prices that are “round” (e.g. \$600, not \$613.12). Further, there is more coordination around “rounder” numbers. That is, even though both \$600 and \$590 are possible strike prices, the strike at \$600 will be a stronger coordination point and have more options issued. For example, Figure 1.4b shows how Sum(Open Interest) and Sum(Gamma*Open Interest) vary with respect to the strike price for options on Apple Inc. (AAPL) on Dec 31, 2012. We can clearly see spikes in the open interest around the “rounder” numbers of \$550, \$600, \$650, \$700, etc.

To use this instrument properly, we need to discard observations where there have been big price movements in the recent past. Specifically, for the two stage least-square (2SLS) estimation, we discard observations where $r_{ma,i,t}^e$ is in the upper and lower 10%. Otherwise, the first stage of the 2SLS estimation is weak for two reasons. First, as the stock price moves away from the strike price, gamma falls to zero (Figure 1.4a). Second, options are generally issued with strike prices near the current stock price (Figure 1.4b). Therefore, after a large

(a) Example: Gamma is Highest Near the Strike Price



(b) Example: Options are Struck at Round Number Prices
On December 31, 2012, Apple Inc. (AAPL) closed at a stock price of \$532.17.



(c) Binscatter on Full Dataset

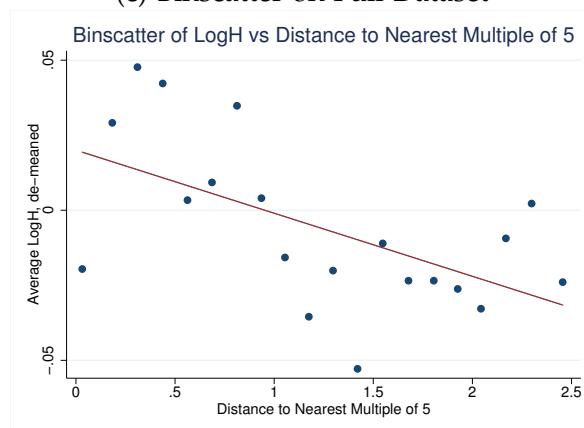


Figure 1.4: *IV, Distance to Nearest Round Number: First Stage Graphs*

Our instrument for hedging demand is the absolute difference between the underlying stock price and the nearest round number. This instrument is based on two facts: (1) put and call gamma is highest near the strike price and (2) as a matter of institutional detail, exchange-traded options are struck at round number prices (e.g. \$600 as opposed to \$613.12).

price increase/decrease, hedging demand is near zero and the instrument is weak. As a result, we discard observations where $r_{ma,i,t}^e$ is in the upper and lower 10%.

Based on a firm's fundamentals, there is no reason to believe that stock prices behave different around round numbers. However, non-fundamental factors could also be a concern. Hence, to verify the exclusion restriction of our instrument, we consider three strands of prior research on stock price behavior near round numbers: (1) Ni, Pearson, and Potesman (2005) document unusual stock price behavior near option strike prices at option expiration. They attribute it either re-hedging or stock price manipulation by option traders. Re-hedging is precisely our mechanism and so that satisfies the desired exclusion restriction. Stock price manipulation likely only happens on the expiration day, since it is extremely costly to manipulate stock prices for more than brief periods of time. In contrast, our instrumental variable regressions focus the effect of a stock price being close to a round number price, in general. (2) Prior research has shown that stock prices cluster at round decimals. For example, Harris (1991) argues that traders focus on discrete price sets to simplify negotiations; Ikenberry and Weston (2007) documents that stock prices cluster at increments of \$0.05 and \$0.10 because traders prefer nickel and dime increments for behavioral reasons.¹³ These findings do not appear to pose an issue to our exclusion restriction. First, it is unclear why such coordination points would create more return autocorrelation. Second, since the average stock price is around \$30, our instrumental variable focuses on much coarser round numbers than those studies, which focus on round decimals. In our instrument, \$30 is a "round" number, whereas \$32.25 is not. (3) Some papers have argued that there are behavioral effects near round number prices of stock indices. For example, Donaldson and Kim (1993) argues that multiples of 100 of the Dow Jones Industrial Average are salient reference points and hence create "price barriers."¹⁴ However, other papers argue that these price barriers are not robust (Ley and Varian, 1994;

¹³See also Christie and Schultz (1994), which famously uncovered that market makers collude by avoiding odd-eighth quotes.

¹⁴See also Pomorski (2008) which examines similar effects in the cross section of stocks.

De Ceuster, Dhaene, and Schatteman, 1998). At the very least, our instrument appears to satisfy the exclusion restriction for the primary confound of news entering option markets first.

To map these institutional details into an instrument, we construct a metric that measures the distance from the stock price to the nearest “round” price. We also need to account for the fact that the distance between successive strike prices rises as with the stock price. For example, for stocks with prices around \$600, open interest clusters around every \$50 (e.g. \$550, \$600, etc.). For stocks with prices around \$60, open interest clusters around every \$5 (e.g. \$55, \$60, etc.). We use the following procedure: First, we normalize prices to have two digits before the decimal point. So, for example \$5.32, \$53.20 and \$532.00 all map to a “normalized price” P_{it}^{norm} of \$53.20. Then, our instrumental variable is the distance of the normalized price to the nearest integer multiple of 5:

$$IV_{it} = Abs(P_{it}^{norm} - \text{NearestMultipleOf5})$$

To illustrate, on December 31, 2012, the close price of AAPL stock is \$532.17. Therefore, the normalized price is \$53.21. Since the nearest multiple of 5 is \$55, so IV_{it} is $1.79 = 55.00 - 53.21$. One can alternately create multiple instrumental variables using the distance of the normalized price to the nearest multiple of \$10 and \$2.5. However, using multiple instruments puts more onus on the first stage not being weak. If the first stage is weak, multiple instruments actually worsens the weak instruments bias (Chapter 4 of Angrist and Pischke, 2009).

Because $LogH$ appears as the main effect ($LogH_{ma,i,t}$) and the interaction effect ($r_{ma,i,t}^e \times LogH_{ma,i,t}$), we have two endogenous variables. Hence, our two instruments are $IV_{ma,i,t}$ and $r_{ma,i,t}^e \times IV_{ma,i,t}$. This implementation avoids the econometrically incorrect “forbidden regression” (see textbooks such as Wooldridge, 2001). It is incorrect to only have *one* first-stage fitted value $\widehat{LogH}_{ma,i,t}$ and use it to construct $r_{ma,i,t}^e \times \widehat{LogH}_{ma,i,t}$ in the second stage. This incorrect procedure produces estimates that are inconsistent, i.e. do not converge to the true parameter. Instead, we must estimate a separate first-stage fitted value for the *entire*

interaction term $\widehat{r_{ma,i,t}^e \times \text{Log}H_{ma,i,t}}$. Intuitively, the heart of the forbidden regression is that if f is a linear projection and g is a non-linear transformation, then $f(g(x)) \neq g(f(x))$. In our situation, the 2SLS estimation is a linear projection and the interaction term is a non-linear transformation.

Our 2SLS estimator is

$$\begin{aligned} r_{i,t+1}^e = & b_{0,t} + b_1 \cdot r_{ma,i,t}^e + \lambda^{IV} \cdot \widehat{r_{ma,i,t}^e \times \text{Log}H_{ma,i,t}} \\ & + b_2 \cdot \widehat{\text{Log}H_{ma,i,t}} + b_3 \cdot r_{ma,i,t}^e \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t} + \epsilon_{i,t+1} \end{aligned} \quad (1.5)$$

Table 1.12 displays the results. As before, we are interested in the coefficient on the interaction term $r_{ma,i,t}^e \times \text{Log}H_{ma,i,t}$, which is $\hat{\lambda}^{IV}$. Because the 2SLS estimator is more comparable to a panel regression than a Fama-MacBeth regression, we compare our 2SLS estimates to a panel regression. Column (1) is a panel regression with fixed effects by time and clustering by time. It is the same as Table 1.8 Column (5). Column (2) is the same panel regression, but with the highest 10% and lowest 10% of $r_{ma,i,t}^e$ discarded. As discussed earlier, this discarding is necessary for instrument strength.

Column (3) shows the results of the 2SLS regression. It is statistically significant and has the same sign the theory in Section 1.2. Combined with the exclusion restriction, this result corroborates the assertion that hedging demand causally increases return autocorrelation. The IV estimate $\hat{\lambda}^{IV}$ is also significantly larger than the panel ordinary least squares estimate $\hat{\lambda}^{panel}$. However, the two estimates are not directly comparable. Variation in hedging demand comes from two sources: the gamma of each option and the number of options being hedged. This instrument focuses on variation due to the former whereas the panel regression focuses on both sources of variation combined.

We also formally test the strength of the instrument in the first stage. As a general econometric matter, while the 2SLS estimator is biased, it is consistent so the estimate converges to the true parameter as the sample size increases. However, when the first stage is weak, asymptotic consistency can be surprisingly slow (Bound, Jaeger, and Baker, 1995). A commonly used rule of thumb is that the F-statistic on the excluded instruments should

Table 1.12: *IV, Distance to Nearest Round Number: 2SLS*

This table displays the results of the two stage least square (2SLS) estimation. Our instrumental variable (IV) is the absolute difference between the price of the underlying stock and the nearest round number; see Section 1.6 for more details. Since the two endogenous variables are $\text{Log}H_{ma,i,t}$ and $r_{ma,i,t}^e \times \text{Log}H_{ma,i,t}$, the two instruments are $IV_{ma,i,t}$ and $r_{ma,i,t}^e \times IV_{ma,i,t}$. This implementation avoids the econometrically incorrect “forbidden regression” (see textbooks such as Wooldridge, 2001). The coefficient of interest is λ^{IV} . Column (1) is a panel OLS regression. It is the same as Table 1.8 Column (5). Column (2) is the same panel regression, but with the highest 10% and lowest 10% of $r_{ma,i,t}^e$ discarded. As we discuss in the main text, this is necessary for instrument strength. Column (3) shows the results of the 2SLS regression. In general, for 2SLS estimation, r-squared can be negative. F-statistic on excluded instruments: 697.7 (Cragg-Donald test), 266.0 (Kleibergen-Paap test).

$$r_{i,t+1}^e = b_{0,t} + b_1 \cdot r_{ma,i,t}^e + \lambda^{IV} \cdot \widehat{r_{ma,i,t}^e \times \text{Log}H_{ma,i,t}} + b_2 \cdot \widehat{\text{Log}H_{ma,i,t}} + b_3 \cdot r_{ma,i,t}^e \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t} + \epsilon_{i,t+1}$$

For legibility, displayed coefficients are regressions estimates multiplied by 100. Controls: log market capitalization, log dollar volume, log share price, and their interactions with returns.

VARIABLES	(1)	(2)	(3)
r_{ma}^e	-3.96*** (-4.4)	-0.56 (-1.1)	-0.41 (-0.7)
$r_{ma}^e \times \text{Log}H_{ma}$	1.35*** (5.5)	1.27*** (5.9)	23.07*** (2.7)
$\text{Log}H_{ma}$	-0.02*** (-7.5)	-0.01*** (-6.7)	-0.00 (-0.1)
Observations	4,326,618	3,456,809	3,456,809
R-squared	0.001	0.000	-0.005
Number of groups	4,434	4,434	4,434
Controls + xterms	Y	Y	Y
StdErr cluster	Time	Time	Time
Fixed effect	Time	Time	Time
Method	Panel OLS	Panel OLS	2SLS

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

exceed 10 (Angrist and Pischke, 2009). Our instrument comfortably passes that test with F-statistics of 266.0 (Kleibergen-Paap test). The F-statistics show that the potential bias due to weak instruments is not a major concern for our setting.

1.7 Conclusion

In the textbook model of derivatives, two end users who want opposite financial exposures come together and make a side bet. For example, a baker and a farmer who use a derivative contract to take opposite positions about wheat prices. Such end-user/end-user trade is a side bet and does not affect the underlying asset. However, when the counterparty is instead a liquidity provider and when the derivative has convex payoffs (as options do), derivatives are not mere side bets.¹⁵ In such situations, the derivative can affect the price dynamics of the underlying asset due to hedging by the liquidity provider. Previous research has shown that net demand by end users can affect option prices (e.g. Green and Figlewski, 1999; Bollen and Whaley, 2004; Garleanu, Pedersen, and Poteshman, 2009). We show how it can affect the underlying prices as well.

We show that hedging by option writers creates an upward sloping demand curve in the underlying stock: to remain hedged, option writers must buy stock after the price rises and sell stock after the price falls. If trading the stock has price impact and dynamic hedging is not instantaneous, then this upward sloping demand curve increases the autocorrelation of stock returns. This mechanism is similar to the forced buying/selling by portfolio insurers, which the Brady Report argues was a significant factor in the October 1987 stock market crash (Brady, 1988).

We develop a measure of hedging demand, which quantifies the sensitivity of the option writers' hedge to changes in the underlying price, and document empirically that stocks with higher hedging demand have higher return autocorrelations. We further provide causal identification using an instrumental variable based on the institutional detail that

¹⁵Purely linear derivatives can also affect the underlying asset through a different channel of relaxing short sale constraints (Banerjee and Graveline, 2014).

exchange-traded options are struck at round number prices. The robustness of the results supports the view that hedging demand indeed increases the return autocorrelation of the underlying stock.

One direct area of future research is to test if options hedging affects other asset classes as well. For example, according to the Bank for International Settlements, the notional amount of interest rate options is roughly ten times the combined notional amount of equity index and individual stock options (BIS, 2014). On the other hand, the interest rates market is more liquid, so dynamic hedging with the underlying asset has less price impact. While many options trade over-the-counter, future research that uses the subset of data on exchange-traded options in other asset classes could likely also apply our instrumental variable (the absolute difference between the underlying price and the nearest round number) for exogenous variation and establish causality.

Another area of future research is to study derivatives other than options. As the upward sloping demand curve comes from the convexity of options, a similar mechanism may apply to hedging other derivatives that are also convex. For example, the theory suggests that credit default swaps could have similar effects. Sellers of credit default swaps often hedge using the stock of the underlying firm and the value of a credit default swap is convex with respect to the underlying stock. Estimates from the Bank for International Settlements (2014) suggest that the market for credit default swaps is also many times the size of the market for equity-related options.

Chapter 2

Disagreement and the Option Stock Volume Ratio

2.1 Introduction

What explains the relationship between trading volume in an options market and trading volume in its underlying market? Prices between the two markets are held together by theories such as Black and Scholes (1973) and Merton (1973). However, those theories are silent about the relationship of trading volume between the two markets. There is a literature that focuses on asymmetric information as an explanation for trading volume in the two markets. These papers emphasize that traders with private information transact in the options market to take advantage of the embedded leverage in options or to alleviate short-sale constraints (Black, 1975; Easley, O'Hara, and Srinivas, 1998). In this paper, I build on the disagreement literature in behavioral finance to propose a disagreement-based model of the trade in an options market and its underlying market. I show how this model helps explain the negative correlation between O/S and the underlying asset's returns in the cross section. I also document other related empirical findings and show how the disagreement model explains those findings better than the asymmetric information model.

My empirical setting focuses on individual stocks and their options. To measure the

relative behavior of trading volume in the options market and in the stock market, I focus on the ratio of options trading volume to stock trading volume, which Roll, Schwartz, and Subrahmanyam (2010) introduce and label as O/S . They discuss the possibility that O/S can be driven by asymmetric information or by disagreement. However, they focus on the time-series properties and determinants of O/S as opposed to its relationship with returns.

Johnson and So (2012) show that O/S is a negative predictor of cross-sectional stock returns: stocks with high O/S underperform stocks with low O/S . Extending the approach of Easley, O'Hara, and Srinivas (1998), they build a model of asymmetric information to explain their findings. In their model, short-sale constraints prevent investors from expressing negative bets in the stock market. They assume that the implied cost of shorting in the options market is lower because options market makers can more easily short stocks than investors. As a result, in their model, O/S is a predictor of negative private information and hence forecasts negative returns in the cross section. While their model has many rational components, Johnson and So (2012) note that their model requires that the investors in the underlying stock be somewhat irrational, since they do not react to the information embedded in options volume.

In contrast, consider a disagreement model, which also predicts a similar negative relationship between O/S and future returns. In this model, investors also face short-sale constraints. However, the fundamental reason that investors trade is because they disagree about the value of an asset, not because one particular set of investors knows more about the value of the asset as in the asymmetric information model. While investors can trade derivatives, the derivatives are not costless and hence do not fully offset the short-sale constraints. In this setting, the stock price becomes over-valued because some of the pessimists are shut out of the market (Miller, 1977). Furthermore, in this model, variation in either belief heterogeneity, risk tolerance, or supply of the underlying asset induces a negative correlation between the quantity of derivatives traded and expected returns, which corresponds to O/S and cross sectional returns in the empirical application. For example, an increase in the belief heterogeneity increases the price of the underlying asset due to the

Miller (1977) effect. It also increases derivatives trade because the optimists and pessimists now have more extreme beliefs. Hence, it induces a negative correlation between derivatives trade and returns.

This disagreement model also produces several other predictions, which I document empirically in this paper and contrast with the predictions of asymmetric information model: First, the ratio of options trading volume to stock trading volume O/S is positively correlated with proxies for disagreement. Following Diether, Malloy, and Scherbina (2002), I use the dispersion in analyst forecasts as a proxy for the amount of belief heterogeneity. I verify that O/S is positively correlated with analyst forecast dispersion. In contrast, I find that O/S is negatively correlated with the bid-ask spread, which is traditionally thought of as a proxy for asymmetric information. This negative correlation is the opposite of the prediction in the asymmetric information model.

Second, the disagreement model can more easily explain the relationship between put volume and call volume and future cross-sectional returns. O/S is the total options trading volume divided by stock trading volume, and we can decompose O/S into a put-volume-version $(O/S)^{put}$ (i.e. put volume divided by stock volume) and a call-volume-version of $(O/S)^{call}$ (i.e. call volume divided by stock volume). In my disagreement model, higher disagreement increases both put volume and call volume and hence forecasts negative returns, which is what I document in the data.

In contrast, the asymmetric information model has potential empirical and theoretical shortcomings with respect to this fact. If the asymmetric information model assumes that investors are long-only, then put option volume should forecast negative returns and call option volume should forecast positive returns. However, this prediction is at odds with the data, which finds that both put option volume and call option volume forecast negative returns in the cross section. One can resolve this empirical problem by instead assuming that investors can also sell options. However, this assumption creates a theoretical tension in the model because it is unclear if an investor who can sell options would also have a significant short-sale constraint, since both involve posting collateral and potentially unlimited losses.

Without the short-sale constraint, the asymmetric information model no longer predicts that O/S forecasts negative returns. By comparison, the disagreement model is consistent with both put and call option volume forecasting negative returns, while still retaining the long-only constraint.

Third, I document empirically that O/S predicts cross-sectional expected returns over a year into the future. This empirical finding is at odds with the asymmetric information model. In the asymmetric information model, the investor with private information expresses his belief through the options market if it is negative news. This model also requires that the underlying stock market not react to trade in the options market. While this assumption is plausible over short time spans, the ability of O/S to predict cross-sectional returns 12+ month into the future is at odds with this assumption of the asymmetric information model. In fact, papers in the asymmetric information literature on options volume often focus on relatively short time spans. For example, Easley, O'Hara, and Srinivas (1998) focus on 30 minute prediction window and, using non-public data, Pan and Poteshman (2006) find predictability on the order of a few weeks. In contrast, in my disagreement model, the derivatives market and the underlying market are in equilibrium. No investor has private information that the other investors are trying to infer. Rather, investors simply agree to disagree with each other.

This long-term predictability is related to, but also distinct from, Johnson and So (2012)'s findings. Their analysis uses data at a weekly frequency and finds that the cross-sectional predictability loses statistical significance at 4-6 weeks. My analysis uses data at a monthly frequency and finds statistical significance at longer time spans. Statistically, my result shows that the lagged values of O/S contain information beyond the latest value of O/S . One possibility is that factors outside my model also affect O/S , which makes the empirical O/S a noisy measure of the theoretical O/S . This noise creates downward bias in the magnitude of the effect and increases the standard errors. If these other factors are random shocks, then aggregating helps dilute the noise and improve the link between the empirical version of O/S and its theoretical counterpart.

Finally, I find strong interaction effects between O/S and analyst forecast dispersion, a proxy for investor disagreement (Diether, Malloy, and Scherbina, 2002). In general, O/S is a negative forecaster of returns in the cross section. However, when analyst forecast dispersion is high, O/S is an even stronger negative forecaster of returns in the cross section. My model suggests there are two competing effects here. In the model, variation in risk tolerance or in quantity of the underlying stock creates a negative correlation between O/S and expected returns. When belief heterogeneity is higher, the cross derivative implies that these factors have a weaker impact on price for each unit change. However, if higher belief heterogeneity is correlated with more variation in these factors, then the interaction effect goes the opposite direction. The empirics suggest that the latter effect is stronger.

The empirical part of my paper connects directly with the recent papers on the option stock volume ratio, O/S . Roll, Schwartz, and Subrahmanyam (2010) introduce this measure and discuss some of its basic properties. Johnson and So (2012) document the relationship between O/S and cross-sectional stock returns. My paper proposes an alternate explanation based on disagreement and explains how the disagreement model helps us better understand other empirical facts about O/S . Ge, Lin, and Pearson (2014) use data on signed options volume to show that the forecasting power of O/S is due to embedded leverage and not due to alleviating short sale constraints. They briefly consider the relationship between O/S and disagreement, but reject it because it does not affect the predictability of O/S at short horizons. However, they note their analysis focuses on weekly frequencies and does not capture the potential effects of slow moving disagreement, which I study here. My paper is also related to other papers on options volume, such as Pan and Poteshman (2006), which uses proprietary data to compute the buyer-initiated put-call volume ratio and analyze its relationship with returns.

The theoretical part of my paper is closely related to the behavioral finance literature on belief disagreement as the reason for trade (Harris and Raviv, 1993; Kandel and Pearson, 1995; Hong and Stein, 2007). My model builds on the foundation of Miller (1977) and Chen, Hong, and Stein (2002) and extends it to allow for derivatives trade. This belief disagreement

model contrasts with the literature on modeling trade in financial markets as the interaction between privately informed investors and noise/liquidity traders (Grossman and Stiglitz, 1980; Kyle, 1985). As emphasized by Hong and Stein (2007), disagreement models can much more easily explain the large trading volumes we observe in financial markets. In asymmetric information models, trade requires liquidity shocks since uninformed but rational investors will infer the private information from market interactions with informed investors.

Section 2.2 describes the disagreement model and its implications. Section 2.3 describes my data sources, defines the key explanatory variables, and discusses summary statistics. Section 2.4 discusses the main empirical results. Section 2.5 concludes.

2.2 A Model of Disagreement and Derivatives Trade

My model builds on the disagreement models of Miller (1977) and Chen, Hong, and Stein (2002) and allows for derivative securities. The model has two time periods $t \in \{0, 1\}$. There are four types of assets: the risk-free asset, the risky asset (“underlying stock”), and two derivatives on the risky asset (a “positive bet” derivative and a “negative bet” derivative). I normalize the net risk-free rate to 0, by assuming that it has elastic supply at that rate. At time 1, each unit of the underlying stock pays a terminal dividend distributed $N(\mu, 1)$. The underlying stock has a total supply of Q_x . For tractability, I assume that both derivative securities are a leveraged payoff of the underlying stock. The “positive bet” derivative security pays n times the terminal dividend of the underlying stock, at time 1. The “negative bet” derivative security pays the opposite, i.e. $-n$ times the terminal dividend of the underlying stock, at time 1. I assume these are separate securities to allow investors with shorting constraints to make negative bets. The “positive bet” derivative and “negative bet” derivative are meant to be analogous to a call option and a put option, respectively. However, because the non-linearity of options creates tractability problems, I instead model derivatives securities that payoff a linear multiple ($\pm n$) of the underlying stock. In terms of quantities, since derivatives are contracts between agents inside the model, there are no

exogenous constraints on the supply of derivatives.

There are two types of agents in the model: derivatives market makers and long-only buyers, each with a unit mass. The derivatives market makers exist only to create the two derivative securities. To prevent them from making other bets in the market, I assume that the derivatives market makers have infinite risk aversion and strive simply to minimize volatility of their wealth. For a cost $n \cdot c > 0$, the derivatives market maker can create a pair of “positive bet” derivatives and “negative bet” derivatives. The market makers assign the statutory incidence of the cost to the “negative bet” derivative. Because derivatives market makers create the derivatives in pairs, they are never exposed to any risk, satisfying their infinite risk aversion.

As their name suggests, long-only buyers can buy the underlying stock or either of the derivatives, but cannot short sell. There are a continuum of long-only buyers with heterogeneous beliefs about the mean payoff of the underlying stock. Long-only buyer i believes that the underlying stock will pay an uncertain terminal dividend with distribution $N(i, 1)$. For simplicity, I assume that they only disagree about the mean of the payoff and not the standard deviation. Let $i \sim \text{Uniform}[\mu - h, \mu + h]$ so the average long-only buyer has the correct belief about the mean payoff μ and the total heterogeneity is $2h$. The long-only buyer has constant relative risk aversion (CARA) utility with risk-tolerance τ over next period's wealth.

Let p_x, p_y, p_z denote the prices of the underlying stock, the “positive bet” derivative, and the “negative bet” derivative, respectively. Given the linearity of the setup, $p_y = n \cdot p_x$ and $p_z = -n \cdot p_x + n \cdot c$, since the derivatives have payoff of $\pm n$ times the underlying asset. In a more generalized (but less analytically tractable) model, investors could trade call options and put options instead of a “positive bet” derivative and a “negative bet” derivative. In this alternate setup, the price of the call option and put options would also be closely linked to the price of the underlying stock; however, the relationship would be non-linear as in most option pricing models.

For long-only buyer i , let x_i denote her demand for the underlying stock, y_i denote her

demand for the “positive bet” derivative, and z_i denote her demand for the “negative bet” derivative. Because long-only buyers cannot short, I assume that $x_i \geq 0$, $y_i \geq 0$, and $z_i \geq 0$ for all i . In equilibrium, the long-only buyers will separate into three groups, $i \in [p_x, \mu + h]$, $i \in [p_x - c, p_x]$, $i \in [\mu - h, p_x]$.

Group $i \in [p_x, \mu + h]$ has an optimistic valuation of the underlying stock and has positive demand for the underlying stock and the “positive bet” derivative. In particular, their demand is $x_i + ny_i = \tau(i - p_x)$. The investors have indeterminate demand between the underlying stock and the “positive bet” derivative because they do not have borrowing constraints. In a more complicated model with borrowing constraints, the most optimistic investors would purchase the “positive bet” derivative for its leverage and the medium optimistic investors would purchase the underlying stock.

Group $i \in [p_x - c, p_x]$ has a moderately pessimistic valuation of the underlying stock and will only hold the risk-free asset. Despite their moderate pessimism, they cannot express their view in the market because the cost of creating a derivative exceeds their modest pessimism.

Group $i \in [\mu - h, p_x - c]$ has a strong pessimistic view of the underlying stock and expresses this negative view by purchasing the “negative bet” derivative. In particular, their demand is $z_i = \frac{\tau((p_x - c) - i)}{n}$ units of the “negative bet” derivative.

Setting supply equal to demand in the underlying stock market and the derivatives market, we get two market clearing conditions:

$$Q_x = \int_{p_x}^{\mu+h} \frac{x_i}{2h} di \quad (2.1)$$

$$\int_{\mu-h}^{p_x-c} \frac{z_i}{2h} di = \int_{p_x}^{\mu+h} \frac{y_i}{2h} di \quad (2.2)$$

As a benchmark, consider the case where $c = 0$. In this case, it is costless to create derivatives, which effectively removes the constraint that buyers must be long-only. In this situation, we have:

$$p_x^{*,bmark} = \mu - \frac{Q_x}{\tau} \quad (2.3)$$

That is, the price of the underlying stock is its true mean payoff less a risk premium. As the

risk-tolerance τ approaches infinity, the risk-premium approaches zero and the price of the underlying stock approaches μ .

Next, I consider the case with general values of c . In this situation, the equilibrium price of the underlying stock is:

$$p_x^* = \mu - \frac{Q_x}{\tau} \cdot \frac{2h}{2h - c} + \frac{c}{2} \quad (2.4)$$

As a regularity condition, we require that:

$$p_x^{*,bmark} - c > \mu - h \quad (2.5)$$

which is equivalent to $h > c + \frac{Q_x}{\tau}$. Intuitively, we need enough belief heterogeneity so the shorting constraint binds.

I define the total quantity of derivatives sold in equilibrium as:

$$Q_z^* := \int_{\mu-h}^{p_x^*-c} \frac{z_i}{2h} di \quad (2.6)$$

Whereas the total supply of the underlying stock Q_x is exogenously fixed, Q_z^* is an endogenous variable since derivatives are synthesized by the market maker. Using this definition, I let Q_z^*/Q_x be the ratio of the quantity of derivatives to the quantity of the underlying stock.

Proposition 4. *The equilibrium price of the underlying stock p_x^* exceeds its benchmark price $p_x^{*,bmark}$.*

This proposition is analogous to the Miller (1977) result. Algebraically, it follows directly from computing the difference $p_x^* - p_x^{*,bmark}$ and imposing the regularity condition in Equation 2.5. Intuitively, the long-only buyers in the group $i \in [p_x^* - c, p_x^*]$ have a negative view of the asset. However, because they cannot short and because of the positive cost of creating the “negative bet” derivative, they are shut out of the market. As a result, p_x^* is higher than the benchmark price $p_x^{*,bmark}$.

Proposition 5. *When c is higher, the quantity of derivatives sold Q_z^* decreases.*

This proposition follows from taking the derivative and applying the regularity condition. Intuitively, when the cost of creating a derivative is higher, it is less attractive to create

derivatives.

Proposition 6. *As belief heterogeneity h increases, the quantity of derivatives sold Q_z^* increases. Since the supply Q_x is held constant, Q_z^*/Q_x (the ratio of the quantity of derivatives sold to the quantity of stock) also decreases.*

The total quantity of “negative bet” derivatives sold by the market maker is Q_z^* (which equals the total quantity of “positive bet” derivatives sold by the market maker). This proposition follows from substituting in p_x^* and applying the regularity condition in Equation 2.5. Intuitively, as the belief heterogeneity increases, the optimists and the pessimists have more divergent beliefs and hence are more interested in purchasing the derivative securities.

Proposition 7. *As belief heterogeneity h increases, the equilibrium price of the underlying stock p_x^* is increases.*

This proposition comes directly from taking the first derivative of p_x^* with respect to h . As shown in Proposition 6, as belief heterogeneity increases, market makers create and sell more derivatives. However, there is still the original supply of Q_x of the underlying stock that the long-only buyers must own in equilibrium. The rise in the equilibrium price of the underlying stock p_x^* is necessary to clear that market.

Proposition 6 and Proposition 7 imply the following corollary:

Corollary 1. *Variation in belief heterogeneity h induces a positive correlation between the quantity of derivatives issued and the price of the underlying stock p_x^* —and hence a negative correlation between the quantity of derivatives sold Q_z^* and the expected return of the underlying stock $\mu - p_x^*$. Similarly, variation in h creates a negative correlation between Q_z^*/Q_x (the ratio of the quantity of derivatives sold to the quantity of stock) and $\mu - p_x^*$ (the expected return of the underlying stock).*

Next, I derive comparative statics with respect to the risk tolerance τ and with respect to the supply of the underlying stock Q_x . In the model, changes in the Q_x are just changes in the supply of the underlying stock. However, we can also think of changes in supply Q_x as coming from changing investor sentiment, as in Chen, Hong, and Stein (2002).

Proposition 8. *As risk tolerance τ increases, p_x^* increases and the quantity of derivatives sold Q_z^* also increases. Since the supply Q_x is held constant, Q_z^*/Q_x (the ratio of the quantity of derivatives sold to the quantity of stock) also decreases.*

For both outcome variables (p_x^* and the quantity of derivatives issued), this proposition follows from taking the first derivative with respect to τ and using the regularity condition in Equation 2.5. Intuitively, as risk tolerance rises, the risk premium decreases, which increases p_x^* . Also, as risk tolerance rises, long-only buyers become more aggressive in their derivative bets on the positive and negative side, which increases the quantity of derivatives issued.

Proposition 9. *As the supply of the underlying stock Q_x increases, p_x^* decreases. Similarly, as the supply Q_x increases, the quantity of derivatives sold Q_z^* decreases and Q_z^*/Q_x (the ratio of the quantity of derivatives issued to the quantity of stock) also decreases.*

Intuitively, as the supply of the underlying stock increases, its price falls as there is now more supply for the buyers to absorb. The quantity of derivatives Q_z^* decreases because more of the optimists in the group $i \in [p_x^*, \mu + h]$ can simply now buy the underlying stock instead of the “positive bet” derivative. Since the quantity of derivatives Q_z^* decreases with an increase in Q_x , the ratio of the quantity of derivatives to the quantity of the underlying stock (i.e. Q_z^*/Q_x) also decreases with an increase in Q_x .

Proposition 8 and Proposition 9 imply the following two corollaries:

Corollary 2. *Variation in the risk tolerance τ or variation in the supply of the underlying stock Q_x induces a positive correlation between the equilibrium quantity of derivatives sold Q_z^* and the price of the underlying stock. This implies a negative correlation between the quantity of derivatives traded and the expected return of the underlying stock $\mu - p_x^*$. Similarly, variation in τ or Q_x creates negative correlation between Q_z^*/Q_x (the ratio of the quantity of derivatives issued to the quantity of stock) and $\mu - p_x^*$ (the expected return of the underlying stock). Because this negative correlation comes from variation in τ and Q_x , this negative correlation persists even if we hold belief heterogeneity constant.*

Corollary 3. *Belief heterogeneity h and Q_z^*/Q_x (the ratio of the quantity of derivatives issued to the quantity of stock) are correlated. However, their correlation will not be 1 since variation in risk tolerance τ or variation in the supply of the underlying stock Q_x also drives Q_z^*/Q_x .*

In the empirics, I focus on the negative correlation between Q_z^*/Q_x (the ratio of the equilibrium quantity of derivatives sold to the quantity of stock) and $\mu - p_x^*$ (the expected return of the underlying stock), as discussed in Corollary 1 and Corollary 2. I do not focus on the unscaled quantity of derivatives issued since that quantity is not comparable across stocks with differing amounts of shares outstanding. I examine both the volume ratio (options trading volume divided by stock trading volume) and the level ratio (using levels as opposed to flows, i.e. open interest in the options divided by shares outstanding in the underlying stock). Because this model only has two periods, there is not a meaningful distinction between the two measures in the model. Finally, in the empirical application, $\mu - p_x^*$ is best thought of as the alpha after adjusting for risk factors.

I also examine how $\frac{\partial p_x^*}{\partial \tau}$ and $\frac{\partial p_x^*}{\partial Q_z}$ changes as h increases:

Proposition 10. *As h increases, $\frac{\partial p_x^*}{\partial \tau}$ declines (i.e. $\frac{\partial p_x^*}{\partial \tau \partial h} < 0$), but $\frac{\partial p_x^*}{\partial \tau}$ is still positive overall. Also, as h increases, $\frac{\partial p_x^*}{\partial Q_x}$ rises (i.e. $\frac{\partial p_x^*}{\partial Q_x \partial h} > 0$), but $\frac{\partial p_x^*}{\partial Q_x}$ is still negative overall.*

This proposition states that when h is larger, each unit increase in τ still increases p_x^* , but at a slower rate (and vice versa for changes in Q_x). This proposition however does not necessarily determine how the relationship between Q_z^*/Q_x and future returns $\mu - p_x^*$ varies with h . For example, if there is more τ variation when h is higher, then we may observe a stronger relationship between Q_z^*/Q_x and future returns $\mu - p_x^*$ for high h , despite the cross derivative.

2.3 Data, Variable Definitions, and Summary Statistics

The option-level data in this paper come from OptionMetrics' Ivy Database. This database covers all U.S. exchange-listed options starting from January 1996. Because this paper focuses on cross-sectional variation, I use data on options on individual stocks, as opposed

to data on options on stock indices. For each option, I pull the data on its volume, its type (i.e. put vs call), and its expiration date. The OptionMetrics data is at a daily frequency and I aggregate to construct monthly measures. I omit firm-month observations for which there is no options volume in the entire month, across all option types.

The stock-level data in this paper come from the Center for Research in Security Prices (CRSP), Standard and Poor's COMPUSTAT, Ken French's Data Library, and Thomson Reuters' Institutional Brokers Estimate System (I/B/E/S). From CRSP, I use the Monthly Stock File to obtain data on monthly holding period returns, volume, close prices, bid and ask prices, etc. From COMPUSTAT, I obtain balance sheet data to compute accounting book value. From I/B/E/S, I obtain data on individual analyst forecasts of future earnings per share.

The main explanatory variable in this paper is the option-stock volume ratio (O/S), as introduced by Roll, Schwartz, and Subrahmanyam (2010) and studied by Johnson and So (2012). While prior work has studied O/S at a weekly frequency, this paper studies it at a monthly frequency. As its name suggests, O/S is defined as the ratio of options volume to stock volume. For a given stock on a given day, there are many different types of options, e.g. put vs call, different strike prices, different expiration dates. To measure "options volume," we sum the volume over all the different types of options. Specifically, let k index all the different type of options. Then, let $OptionVolume_{i,t,k}$ denote the volume of an option of type k on a given stock i at time t . For example, on December 31, 2012, there were 3904 contracts traded for the put option on Apple, Inc. stock with an expiration date of February 16, 2013 and a strike price of \$600. I define O/S as:

$$(O/S)_{i,t} := \frac{\sum_k 100 \cdot OptionVolume_{i,t,k}}{StockVolume_{i,t}} \quad (2.7)$$

I multiply $OptionVolume_{i,t,k}$ by 100 because each options contract corresponds to 100 shares of the underlying stock. I also study some specialized versions of O/S defined over specific subsets of options. For example, O/S where options volume is summed only over put options or only over call options. In each of these instances (e.g. Table 2.5), I denote the

subset used.

I also study *LevelO/S*, which defines *O/S* in terms of levels instead of flows/volumes:

$$(LevelO/S)_{i,t} := \frac{\sum_k 100 \cdot OptionOpenInterest_{i,t,k}}{StockSharesOutstanding_{i,t}} \quad (2.8)$$

I measure investor disagreement using the dispersion in analyst forecasts (*Disp*), as used in Diether, Malloy, and Scherbina (2002). From the I/B/E/S Unadjusted Detail file, I pull the individual analyst forecasts of earnings per share for the next fiscal year. As highlighted by Diether, Malloy, and Scherbina (2002), it is important to use the original forecasts, unadjusted for share splits, due to rounding issues in I/B/E/S. For stock i at time t , let $forecast_{i,t,j}$ denote the forecast by analyst j . Then, *Disp* is defined as

$$Disp_{i,t} := \frac{\sigma(forecast_{i,t,j})}{|E[forecast_{i,t,j}]|} \quad (2.9)$$

For ease of reference, in Table 2.1, I list the key variables in this paper and their definitions. Table 2.2a lists the summary statistics for the full dataset. In the time series and the cross section, the main constraint is the OptionMetrics database. In terms of time span, my dataset covers 1996 to 2013 for 211 months of observations. In terms of the cross section, my dataset has roughly 1500 firms per cross section. Since my dataset omits stocks without any options, the average stock in my dataset has a market capitalization of roughly \$6 billion.

Table 2.2b displays the mean of selected variables, across the quintiles of *O/S*. Stocks with higher *O/S* are have larger market capitalizations, lower book to market ratios, and are more liquid (higher turnover, lower spread). Stocks with higher *O/S* also have higher dispersion in analyst forecasts and higher stock return volatility.

2.4 Main Empirical Results

In this section, I compare and contrast the predictions of the model in Section 2.2 with the data. A key quantity in my model is Q_z^*/Q_x (the ratio of the quantity of derivatives issued to the quantity of stock). My model is a two period model, so taken literally, the model does not distinguish between volume and shares outstanding. As a result, in the empirics, I

Table 2.1: List of Key Variables

This table lists the key variables used in this paper.

Variable	Symbol	Description
Option Stock Volume Ratio	O/S	Options volume (summed over all strike prices, expiration dates, and put/calls) divided by stock volume.
Level Option Stock Ratio	$LevelO/S$	This variable is the analogy of O/S , but in levels. It is the total options open interest divided by stock shares outstanding.
Quintile of Variable X	$Qtile(X)$	Cross-sectional quintile of variable X.
Dispersion of Analyst Forecasts	$Disp$	Standard deviation of analyst forecasts divided by the absolute value of the mean forecast. That is, $\sigma(forecast)/ E[forecast] $. Prior research has used this variable as a proxy for investor disagreement.
Bid-Ask Spread	$Spread$	(Ask - Bid)/(Midpoint) of the underlying stock.
Stock Return Volatility	Vol	Volatility of stock return over the past 60 trading days.
Simple Return	R	Net simple returns, including dividends.
Excess Return	R^e	Return in excess of the one-month Treasury bill.
Market Cap	$MktCap$	Total equity market capitalization of a firm, using close prices.
CAPM Beta	β	I estimate betas using the Scholes and Williams (1977) method. I compute the betas quarterly and use the estimates from the previous quarter to avoid look-forward bias.
Book/Market Ratio	B/M	(Book Equity)/(Market Capitalization). I lag by two quarters to ensure that the accounting data was already publicly available on each date. Following Fama and French (1993), I define Book Equity = Stockholder's Equity + Deferred Taxes and Investment Tax Credit (if available) - Preferred Stock.
Turnover	$Turnover$	(Share Volume)/(Shares Outstanding)

Table 2.2: Summary Statistics

This table displays summary statistics of the main variables. Panel (a) displays summary statistics for the full dataset. Panel (b) displays the mean of selected variables by quintile of O/S . Panel (c) displays the mean of selected variables by quintile of $Disp$. O/S is the ratio of option trading volume to stock trading volume. $LevelO/S$ is the level version of O/S ; it is the total option open interest divided by stock shares outstanding. $MktCap$ is the market capitalization of the underlying stock in billions of dollars. B/M is the book to market ratio. $Turnover$ is the monthly turnover in the underlying stock. $Spread$ is the bid-ask spread of the underlying stock, expressed as a percent. Vol is the daily return volatility over the last 60 trading days. $Disp$ is the dispersion in analyst forecasts, as defined in Diether, Malloy, and Scherbina (2002). See Table 2.1 for full list of variable definitions. Data frequency is monthly, spanning 1996 to 2013, and has $N = 300,860$ firm-month observations.

(a) Full Dataset					
	mean	sd	p10	p50	p90
O/S, in pct	7.57	14.3	0.44	3.65	18.0
LevelO/S, in pct	48.5	97.7	1.91	16.8	115.7
MktCap, in billions	5.82	19.3	0.28	1.37	11.7
B/M Ratio	0.52	0.46	0.14	0.42	0.98
Monthly Turnover, in pct	21.6	19.8	5.32	15.6	44.7
Spread, in pct	0.61	0.99	0.026	0.16	1.80
Num Firms per Month	1590.1	211.1	1296	1592	1871
Vol, in pct	2.76	1.45	1.32	2.42	4.64

(b) Mean of Selected Variables, by quintile of O/S					
	1	2	3	4	5
O/S, in pct	0.47	1.63	3.64	7.70	23.87
LevelO/S, in pct	4.38	11.76	24.44	51.68	146.97
MktCap, in billions	1.75	2.41	3.33	6.05	15.22
B/M Ratio	0.63	0.56	0.52	0.47	0.42
Monthly Turnover, in pct	13.87	17.05	20.60	24.95	30.90
Spread, in pct	0.68	0.65	0.60	0.56	0.55
Vol, in pct	2.45	2.66	2.82	2.92	2.94
Quintile of Disp	2.88	2.92	2.98	3.02	3.13

mean coefficients

(c) Mean of Selected Variables, by quintile of $Disp$					
	1	2	3	4	5
StdDev of O/S, in pct	9.23	11.03	12.84	11.72	11.23
StdDev of LvlO/S, in pct	54.32	67.08	84.63	99.81	125.87

mean coefficients

Table 2.3: Correlation Table of Quintiles of Selected Variables

This table shows the correlation between the quintiles of selected variables. O/S is the ratio of option trading volume to stock trading volume. $LevelO/S$ is the level version of O/S ; it is the total option open interest divided by stock shares outstanding. $Disp$ is the dispersion in analyst forecasts, as defined in Diether, Malloy, and Scherbina (2002). $Spread$ is the bid-ask spread of the underlying stock, expressed as a percent. See Table 2.1 for full list of variable definitions. Data frequency is monthly, spanning 1996 to 2013, and has $N = 300,860$ firm-month observations.

	Qtile(O/S)	Qtile($LevelO/S$)	Qtile($Disp$)	Qtile($Spread$)
Qtile(O/S)	1			
Qtile($LevelO/S$)	0.789***	1		
Qtile($Disp$)	0.0603***	0.153***	1	
Qtile($Spread$)	-0.155***	-0.163***	0.174***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

examine both the flow version O/S and the level version $LevelO/S$. Since prior work in Roll, Schwartz, and Subrahmanyam (2010) and Johnson and So (2012) focus on O/S , I take that as the main baseline and examine $LevelO/S$ as an alternate definition. Where appropriate, I also discuss how the empirical results help us distinguish between the belief disagreement model and the asymmetric information model in Johnson and So (2012).

2.4.1 Correlations with O/S

Corollary 3 predicts that belief heterogeneity should be correlated with O/S and $LevelO/S$. Following Diether, Malloy, and Scherbina (2002), I proxy belief heterogeneity using the dispersion in analyst forecasts $Disp$. The correlation in Table 2.3 verifies this prediction. Both O/S and $LevelO/S$ have a positive and statistically significant correlation with $Disp$. In the model, this correlation comes from the fact that changes in belief heterogeneity cause changes in the quantity of derivatives traded. Furthermore, while the correlation is statistically significant, it is much less than 1. This is also consistent with Corollary 3. In my model, changes in risk tolerance and changes in the total supply of the underlying stock also affect the quantity of derivatives. As noted in Section 2.2, changes in the supply of the

underlying stock can be interpreted as literal changes in supply or also changes in investor sentiment. Because these other factors also affect the quantity of derivatives, it limits the strength of the correlation between *Disp* and *O/S* (or *LevelO/S*).

Table 2.3 also displays correlations that are contrary to the asymmetric information model of options trade. In the table, we see that higher bid-ask spreads in the underlying stock *Spread* are associated with lower values of *O/S* and *LevelO/S*. However, bid-ask spreads are often associated with asymmetric information (Glosten and Milgrom, 1985). Hence, the data suggest that higher *O/S* is associated with lower asymmetric information, which is contrary to the asymmetric information model.

2.4.2 Long-Short Portfolios

Table 2.4 displays the alphas of long-short portfolios. I form the long-short portfolios by sorting stocks into quintiles using the previous month's *O/S* (Columns (1) - (3)) or *LevelO/S* (Columns (4) - (6)). Let $R_{Q5,t+1} - R_{Q1,t+1}$ denote the equal-weighted return spread between the highest quintile and the lowest quintile. I then run the time series regression

$$R_{Q5,t+1} - R_{Q1,t+1} = \alpha + \beta' \cdot f_{t+1} + \epsilon_{t+1} \quad (2.10)$$

where f_{t+1} is a vector of return factors. I initially control for the Carhart (1997) four-factor model, which includes the market factor $R_m - R_f$, the size factor *SMB*, the value factor *HML*, and the momentum factor *UMD* (Column (1) and (4)). Across the different specifications, there is a significant negative loading on the *HML* factor. This loading is consistent with the summary statistics in Table 2.2, which shows that the book-to-market ratio increases with the quintile of *O/S*. I then add the short-term reversal factor *STRev* and the long-term reversal factor *LTRev* (Columns (2) and (5)). Finally, I add the Pastor and Stambaugh (2003) liquidity factor *LIQ* and the Frazzini and Pedersen (2014) betting against beta factor *BAB* (Columns (3) and (6)). Since Table 2.2 shows that *O/S* is correlated with liquidity in the underlying stock (higher turnover, lower spreads), I control for *LIQ* to check that the liquidity factor does not drive the *O/S* return spread. I control for *BAB* because

the long-short portfolio has a low loading on the market factor, so it is important to verify that *BAB* does not drive the *O/S* return spread.

Across the various specifications, several patterns emerge. First, the alphas are all consistently negative with a coefficient size of about -1% per month and a t-statistic in excess of 5. Harvey, Liu, and Zhu (2014) stress that t-statistics for cross-sectional expected return tests should exceed 3 in absolute value, due to data mining concerns. All the specifications here pass that test. Second, we see that the additional controls beyond the Carhart (1997) four-factor model have minimal loadings. For this reason, we use the Carhart (1997) four-factor model as the basic set of controls in later analyses. The findings in this table are consistent with the prior work of Johnson and So (2012). They focus on weekly returns, whereas I focus on monthly returns. Another minor difference is that this table shows the predictability of a new variable *LevelO/S*, which they do not consider.

Table 2.5 examines the effect of forming *O/S* with different subsamples. Recall that *O/S* is defined as (Total Trading Volume of Across All Options)/(Trading Volume of the Underlying Stock). However, we can also define alternate versions of *O/S* only using a subsample of options. For example, we could define $(O/S)^{put}$ where the numerator is instead Total Trading Volume of Put Options. Table 2.5 examines the results of forming alternate versions of *O/S* using different subsamples. Columns (1) defines *O/S* using only put option volume. Column (2) uses only call option volume. Column (3) uses only near-dated options, which I define as options expiring within 30 days. Column (4) uses only far-dated options, which I define as options expiring after 30 days. Since Table 2.4 showed that return factors beyond the Carhart (1997) four-factor model do not have significant loadings, I use the four-factor model as a control. In all subsamples, we continue to see statistically significant negative alphas with t-statistics exceeding 4.5 in absolute value.

Having described the basic features of Table 2.4 and Table 2.5, I now relate it to the model in Section 2.2. The negative alphas are consistent with the predictions from Corollary 1 and Corollary 2. In my belief disagreement model, variation in belief heterogeneity, risk tolerance, or supply shocks, all induce a negative correlation between *O/S* and returns.

Table 2.4: Alpha of a Long-Short Portfolio

This table shows the alphas of long-short portfolios. In Columns (1) - (3), the sorting variable is O/S , the ratio of the total trading volume of options to the trading volume of the underlying stock. In Columns (4) - (6), the sorting variable is $LevelO/S$, the ratio of the total open interest of options to the shares outstanding of the underlying stock. I sort stocks into quintiles based on the previous month's O/S (or $LevelO/S$). The dependent variable is the monthly return spread between the highest quintile and the lowest quintile, $R_{Q5,t+1} - R_{Q1,t+1}$. Portfolios are equal-weighted. Columns (1) and (4) controls for the Carhart (1997) four-factor model. Columns (2) and (5) adds controls for short-term reversal factor (ST Rev) and long-term reversal factor (LT Rev). Columns (3) and (6) adds controls for the Pastor and Stambaugh (2003) liquidity factor (LIQ) and the Frazzini and Pedersen (2014) betting against beta factor (BAB). Time span is monthly data from 1996 to 2013 for a total of 211 months. In parentheses, I display Newey-West t-statistics with lag length 12 months.

$$R_{Q5,t+1} - R_{Q1,t+1} = \alpha + \beta' \cdot f_{t+1} + \epsilon_{t+1}$$

VARIABLES	(1) O/S	(2) O/S	(3) O/S	(4) LevelO/S	(5) LevelO/S	(6) LevelO/S
Rm-Rf, in %	0.25*** (5.7)	0.24*** (5.3)	0.24*** (5.0)	0.34*** (6.6)	0.33*** (6.1)	0.32*** (6.0)
SMB, in %	-0.04 (-0.6)	-0.01 (-0.2)	-0.01 (-0.2)	0.17*** (3.2)	0.19*** (2.9)	0.19*** (2.8)
HML, in %	-0.56*** (-10.9)	-0.53*** (-7.9)	-0.52*** (-7.1)	-0.71*** (-9.5)	-0.69*** (-7.5)	-0.63*** (-6.2)
UMD, in %	0.08* (1.8)	0.09* (1.7)	0.09* (1.7)	-0.05* (-1.7)	-0.04 (-1.1)	-0.03 (-0.7)
ST Rev, in %		0.03 (0.8)	0.03 (0.7)		0.06 (1.3)	0.05 (1.2)
LT Rev, in %		-0.07 (-1.0)	-0.06 (-0.9)		-0.06 (-0.7)	-0.08 (-0.9)
LIQ, in %			0.02 (0.4)			0.02 (0.5)
BAB, in %			-0.01 (-0.2)			-0.07 (-1.5)
Constant	-0.96*** (-5.5)	-0.97*** (-5.3)	-0.97*** (-5.3)	-1.02*** (-5.8)	-1.03*** (-5.6)	-1.01*** (-5.3)
Observations	211	211	211	211	211	211
R-squared	0.584	0.588	0.588	0.696	0.699	0.703

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 2.5: *Alpha of a Long-Short Portfolio: O/S, Different Subsamples*

This table shows the effect of defining *O/S* with different subsamples of the options data. Column (1) is *O/S* defined only using volume of put options. Column (2) only uses volume of call options. Column (3) only uses volume of near-dated options, i.e. expiring within 30 calendar days. Column (4) only uses volume of far-dated options, i.e. expiring after 30 calendar days. For the portfolio sort, I sort stocks into quintiles based on the previous month's *O/S* (defined using different subsamples of options volume). The dependent variable is the monthly return spread between the highest quintile and the lowest quintile, $R_{Q5,t+1} - R_{Q1,t+1}$. Portfolios are equal-weighted. I control for the return factors of the Carhart (1997) four-factor model. Time span is monthly data from 1996 to 2013 for a total of 211 months. In parentheses, I display Newey-West t-statistics with lag length 12 months.

$$R_{Q5,t+1} - R_{Q1,t+1} = \alpha + \beta' \cdot f_{t+1} + \epsilon_{t+1}$$

VARIABLES	(1) Put	(2) Call	(3) Near	(4) Far
Rm-Rf, in %	0.23*** (5.9)	0.24*** (5.4)	0.29*** (5.7)	0.20*** (5.7)
SMB, in %	-0.08 (-1.5)	-0.04 (-0.6)	-0.02 (-0.3)	-0.06 (-1.0)
HML, in %	-0.50*** (-11.2)	-0.57*** (-9.7)	-0.69*** (-9.1)	-0.42*** (-9.8)
UMD, in %	0.01 (0.3)	0.12*** (2.7)	0.07 (1.5)	0.08* (1.7)
Constant	-0.91*** (-6.2)	-0.92*** (-4.8)	-1.10*** (-5.3)	-0.77*** (-4.8)
Observations	211	211	211	211
R-squared	0.566	0.585	0.659	0.470

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Furthermore, in the model, trade in both the “positive bet” derivative (the model’s analogy of call options) and the “negative bet” derivative (the model’s analogy of put options) is associated with negative returns. This is consistent with the empirical evidence that $(O/S)^{put}$ and $(O/S)^{call}$ forecast negative returns in the cross section.

In contrast, an asymmetric information model of options trade has empirical and theoretical tensions regarding the relationship between put/call volume and returns. If the informed traders in the asymmetric information model are long-only, then put option volume should forecast negative returns and call option volume should forecast positive returns. However, Table 2.5 shows that this prediction is counterfactual.

One can attempt to fix the asymmetric information model by allowing informed traders to both buy and sell options. This fix allows the informed trader to act on negative news by selling call options and hence potentially predicting that $(O/S)^{call}$ forecasts negative returns. While this model resolves the empirical issue, it creates a theoretical tension because it is unclear if an investor who can sell options would still have a significant short-sale constraint. In the asymmetric information model, the short-sale constraint is the key reason that an informed trader with negative information chooses to trade in the options market, instead of the stock market. Without the short-sale constraint, the asymmetric information model does not predict that O/S forecasts negative returns in the cross section. In contrast, my disagreement model can explain the reason both put and call option volume forecast negative returns, while still retaining the long-only constraint.

Table 2.6 displays the results of a Fama and MacBeth (1973) regression on the cross section of individual stock returns. This analysis is similar to the portfolio sorting analysis from above, with some small differences. The Fama-MacBeth methodology allows us to control for individual firm characteristics, as opposed to covariances with return factors such as SMB or HML (Daniel and Titman, 1997). However, one drawback of the Fama-MacBeth methodology is that it is sensitive to outliers and regression misspecification. Hence, it is standard to analyze cross-sectional predictability using both portfolio sorts and Fama-MacBeth regressions, as I do here.

Table 2.6 runs the following Fama-MacBeth regression

$$R_{i,t+1}^e = b_{0,t} + \lambda \cdot (\text{Measure of } O/S)_{i,t} + b_1 \cdot X_{i,t} + \epsilon_{i,t+1} \quad (2.11)$$

The dependent variable is the excess monthly return of a stock $R_{i,t+1}^e$, which is its monthly return less the one month Treasury bill rate. For controls $X_{i,t}$, I use the firm characteristics of log market capitalization, log book to market ratio, the CAPM beta, returns over the past 12 months (excluding the most recent month), returns over the past month, and turnover.

Across the different columns, I use variants of O/S . Column (1) uses the quintile of O/S and Column (2) uses the quintile of $LevelO/S$. In both of these regressions, the Fama-MacBeth coefficient is roughly $\lambda = -0.13$ with a t-statistic exceeding 3 in magnitude. The coefficient of -0.13 corresponds to a return difference between the highest quintile and low quintile of $-0.52\% = -0.13\% \times (5 - 1)$ per month, which is smaller than the portfolio sort alpha of roughly -1% in Table 2.4.

Column (3) uses O/S itself and Column (4) uses $LevelO/S$ itself. These regressions differ from the regressions in Column (1) and Column (2) in that they allow the Fama-MacBeth regression to use variation in the dispersion of O/S or $LevelO/S$. Column (1) and Column (2) suppress such information because it uses quintiles of O/S or $LevelO/S$, which forces a constant dispersion over time. In Column (3), the coefficient on O/S is -1.7 (with $|t| > 3$). If we take the average value of O/S in the highest quintile and the lowest quintile, the regression in Column (3) predicts a high-low-quintile return spread of -0.32% per month. In Column (4), the coefficient on $LevelO/S$ is -0.40 (with $|t| > 3$), which predicts a high-low-quintile return spread of -0.48% per month.

2.4.3 Effect of Changing the Portfolio Formation Period

In the previous section, I sorted stocks into portfolios using the previous month's O/S quintiles. In this section, I examine the effect changing the portfolio formation period. Consider sorting stocks into portfolios using the O/S quintiles from j months ago. For example, $j = 1$ would correspond to the sorting methodology from the previous section.

Table 2.6: *Fama-MacBeth Regression: Cross Section of Individual Stock Returns*

This table displays the results of Fama and MacBeth (1973) regressions of excess monthly returns of an individual stock on O/S (and its variants) and on firm characteristics. Excess monthly returns $R_{i,t+1}^e$ are a stock's monthly return less the one month Treasury bill rate. In each regression, I control for the firm characteristics of log market capitalization, log book to market ratio, the CAPM beta, returns over the past 12 months (excluding the most recent month), returns over the past month, and turnover. Column (1) uses the quintile of O/S . Column (2) uses the quintile of $LevelO/S$. Column (3) uses O/S itself. Column (4) uses $LevelO/S$ itself. Data frequency is monthly, spanning 1996 to 2013, and has $N = 300,860$ firm-month observations. In parentheses, I display Newey-West t-statistics with 12 months of lags.

$$R_{i,t+1}^e = b_{0,t} + \lambda \cdot (\text{Measure of } O/S)_{i,t} + b_1 \cdot X_{i,t} + \epsilon_{i,t+1}$$

	(1)	(2)	(3)	(4)
VARIABLES				
Qtile(O/S)	-0.14*** (-3.3)			
Qtile($LevelO/S$)		-0.12*** (-3.1)		
O/S			-1.77*** (-3.2)	
$LevelO/S$				-0.40*** (-3.1)
$\ln(\text{MktCap})$	0.02 (0.4)	0.00 (0.0)	0.01 (0.1)	-0.01 (-0.1)
$\ln(B/M)$	0.14* (1.8)	0.15* (1.9)	0.15* (2.0)	0.16** (2.0)
Beta	-0.14 (-0.9)	-0.13 (-0.9)	-0.15 (-1.0)	-0.14 (-1.0)
$R_{-12mo:-1mo}^i$	-0.00 (-0.3)	-0.00 (-0.4)	-0.00 (-0.4)	-0.00 (-0.4)
$R_{-1mo:-1mo}^i$	-0.02*** (-3.1)	-0.02*** (-3.2)	-0.02*** (-3.1)	-0.02*** (-3.2)
Turnover	-0.65 (-1.6)	-0.55 (-1.4)	-0.77* (-1.7)	-0.46 (-1.0)
Observations	300,860	300,860	300,860	300,860
R-squared	0.073	0.073	0.072	0.072
Number of months	211	211	211	211

t-statistics in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Let $R_{Q5,t+1}^{(j)} - R_{Q1,t+1}^{(j)}$ denote the equal-weighted return spread formed using O/S from j periods ago, i.e. $(O/S)_{i,t-j}$. I run the time-series regression

$$R_{Q5,t+1}^{(j)} - R_{Q1,t+1}^{(j)} = \alpha^{(j)} + \beta' \cdot f_{t+1} + \epsilon_{t+1} \quad (2.12)$$

For the return factors f_{t+1} , I use the Carhart (1997) four-factor model, since Table 2.4 shows that return factors beyond the four-factor model do not have significant loadings. While I examine the effects of changing the lag length of the portfolio formation, I keep the portfolio holding period to be one month, so the dependent variable is always the one month return spread between the quintiles. Figure 2.1 plots the alphas $\alpha^{(j)}$ against the lag length of j months. For both O/S and $LevelO/S$, there is significant persistence in the cross-sectional predictability. For both variables, using the O/S (or $LevelO/S$) value today can predict cross-sectional returns over a year into the future.

In contrast, Johnson and So (2012) only find predictability up to 4-6 weeks into the future. Empirically, the reason for the differing findings is the frequency of the data. They use data at a weekly frequency and I use data at a monthly frequency. The fact that I find longer term cross-sectional predictability suggests that the lagged values of O/S contain information beyond the latest value of O/S . In undisplayed results, I check that forecasting next week's returns using a lagged moving average O/S yields larger estimates and t-statistics than only using last week's O/S .

Measurement error is one possible reason that using monthly O/S has more statistical power than weekly O/S . In my model, variation in both belief heterogeneity and risk tolerance creates a negative relationship between the theoretical version of O/S and future returns. However, in the data, O/S can vary due to other factors. If these other factors are random shocks, then weekly O/S has more measurement error than monthly O/S (relative to the theoretical counterpart in my model). This measurement error creates a downward bias in the magnitude of the effect and also reduces the statistical significance.

An alternate possibility is that aggregating the dependent variable (returns) increases the statistical significance. I do not believe that is the underlying reason for the difference

between my results and Johnson and So (2012), but I explain this possibility for completeness. In the time-series return predictability literature, long horizon return regressions on dividend yield have larger coefficients and stronger statistical significance than short horizon regressions. Campbell (2001) and Cochrane (2007) explain that this finding is because (1) dividend yields are persistent and (2) return shocks and dividend yield shocks are negatively correlated, due to the Campbell and Shiller (1988) decomposition. One would need to first adapt those theories to the portfolio sorting test methodology that cross-sectional asset pricing papers use. Furthermore, in the case of O/S and its quintiles, the regressor has much less persistence and there is not necessarily a clear reason for a negative relationship between shocks to returns and shocks to the regressor.

Panel (b) illustrates this result in a different way, using $LevelO/S$, which is a version of O/S not previously examined in the literature. $LevelO/S$ is the ratio of options open interest and stock shares outstanding (as opposed to options volume and stock volume, which is a flow variable). Since $LevelO/S$ is defined using a level as opposed to flow measure of quantity, it is less subject to weekly variation. Panel (b) shows that $LevelO/S$ has slightly larger coefficients and forecasts returns at a slightly longer time spans than O/S itself.

The results in Figure 2.1 can also help us distinguish between the asymmetric information model and the belief heterogeneity model. In the asymmetric information model, investors with private negative information express their negative beliefs in the options market due to short-sale constraints. However, in this model, it is important that the underlying stock price does not react to options trade. Over short time spans, this assumption seems reasonable. However, the empirical analysis in Figure 2.1 shows that O/S can predict cross-sectional returns over a year later, which creates empirical tension for the asymmetric information model.

In contrast, the belief heterogeneity model accommodates the long term predictability more easily. In the belief heterogeneity model, increases in belief heterogeneity drive optimists and pessimists to purchase more derivatives. At the same time, belief heterogeneity increases the underlying price due to short-sale constraints. In this model, the prices in the

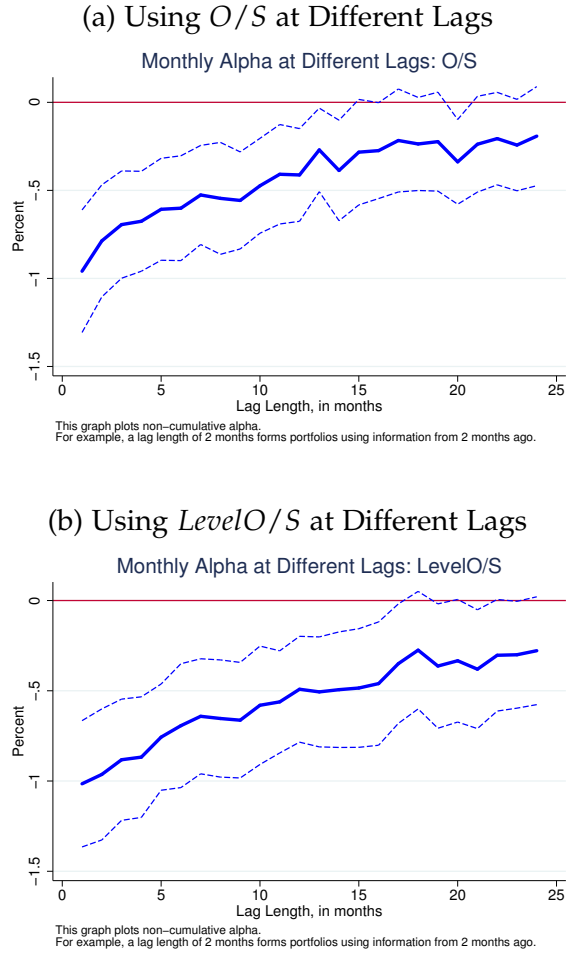


Figure 2.1: Monthly Alpha at Different Lags

These figures show the effect of changing the lag length of the sorting variable on portfolio alpha, controlling for the Carhart (1997) four-factor model. In Panel (a), the sorting variable is O/S . In Panel (b), the sorting variable is $LevelO/S$. The portfolios are equal weighted monthly returns for the highest quintile less the lowest quintile of the lagged sorting variable. I plot the non-cumulative alpha, i.e. the portfolio holding period is always one month. Formally, suppose I use the quintiles of O/S from j months ago to sort stocks into quintiles. Let $R_{Q5,t+1}^{(j)} - R_{Q1,t+1}^{(j)}$ denote the high-low return spread of this portfolio. I run the regression

$$R_{Q5,t+1}^{(j)} - R_{Q1,t+1}^{(j)} = \alpha^{(j)} + \beta' \cdot f_{t+1} + \epsilon_{t+1}$$

and plot the values of $\alpha^{(j)}$. The dotted lines show the 95% confidence interval, computed using Newey-West standard errors with 12 months of lags. The time span is monthly data from 1996 to 2013 for a total of 211 months.

derivatives market and the underlying market are in equilibrium (though, as with all disagreement models, the equilibrium is because the investors agree to disagree). Furthermore, it seems plausible that disagreement in beliefs between investors in the market can persist over a year.

2.4.4 Interaction Effects with Analyst Forecast Dispersion

Here I examine the interaction effect between O/S and analyst forecast dispersion $Disp$. The dispersion in analyst forecasts is often used as a proxy for belief disagreement between investors. Diether, Malloy, and Scherbina (2002) show that stocks with higher analyst forecast dispersion have lower returns, which they argue is evidence supporting the belief disagreement model of Miller (1977). For stock i at time t , let $forecast_{i,t,j}$ denote the forecast by analyst j . Then, $Disp$ is defined as¹

$$Disp_{i,t} := \frac{\sigma(forecast_{i,t,j})}{|E[forecast_{i,t,j}]|}$$

Since this analysis requires data on analyst forecasts, the dataset in this subsection is a subset of the dataset in the rest of this paper because there are stocks with options, but no analyst coverage. The dataset spans also 1996 to 2013, but has roughly 30% fewer firm-month observations ($N = 233,049$).

Table 2.7a shows the average excess monthly returns of a two way portfolio sort between the quintiles of O/S vs the quintiles of analyst forecast dispersion $Disp$. As we move down the table, the quintile of O/S increases. As we move the right of the table, $Disp$ increases. The bottom row of Table 2.7a displays the average monthly excess return of the high-low-quintile return spread of O/S , across the different quintiles of $Disp$. For example, if we fix $Qtile(Disp) = 1$, the average monthly excess return of $Qtile(O/S) = 5$ minus the average monthly excess return of $Qtile(O/S) = 1$ is -0.57% per month. Furthermore, we

¹Some papers also use an alternate definition of $Disp$ where the denominator is the unadjusted share price, instead of the mean analyst forecast of earnings per share. This alternate definition can help potentially avoid divide by zero problems when the earnings per share is very low. Though not displayed, I have also computed this alternate definition and found similar results.

observe that the size of this return spread is increasing with $Qtile(Disp)$. That is, there is an interaction effect between O/S and $Disp$.

While the two way portfolio sort in Table 2.7a is a convenient non-parametric way to look at the pattern between O/S and $Disp$, one might wonder if the interaction effect persists after controlling for other variables. To assess whether the interaction effect survives the addition of controls, Table 2.7b uses Fama-MacBeth regressions on the cross section of individual stock returns. I run the regression

$$\begin{aligned} R_{i,t+1}^e = & b_{0,t} + \lambda_1 \cdot Qtile(O/S)_{i,t} + \lambda_2 \cdot Qtile(O/S)_{i,t} \times Qtile(Disp)_{i,t} \\ & + \lambda_3 \cdot Qtile(Disp)_{i,t} + b_1 \cdot X_{i,t} + \epsilon_{i,t+1} \end{aligned} \quad (2.13)$$

The dependent variable is excess monthly return of a given stock i in a given month $t + 1$. The controls are the same as the controls in Table 2.6, i.e. log market capitalization, log book to market ratio, the CAPM beta, returns over the past 12 months (excluding the most recent month), returns over the past month, and turnover.

Because this table also examines an interaction effect, I cross-sectionally de-mean the controls and the quintiles. For the terms uninvolved in the interaction, this de-meaning does not affect the regression coefficient, since the Fama-MacBeth regression allows for a different intercept for each time period. However, it is important to cross-sectionally de-mean the variables $Qtile(O/S)$ and $Qtile(Disp)$. Otherwise, in the regression with the interaction term, the coefficient λ_1 on $Qtile(O/S)$ will correspond to the effect when $Qtile(Disp) = 0$, which is undefined as quintiles range from 1 to 5. By cross-sectionally de-meaning, the coefficient λ_1 on $Qtile(O/S)$ corresponds to the effect for a firm with an average value of $Qtile(Disp)$, which is 3.

Column (1) of this table is analogous to Column (1) of Table 2.6, but for the subset of data with analyst forecasts. I verify that the coefficients are essentially the same, confirming that this dataset is a reasonably representative sample of the dataset used in the rest of this paper. In Column (2), I additionally control for the quintile of $Disp$. It only marginally affects the coefficient on the quintile of O/S . Over this time interval, $Disp$ is not statistically

Table 2.7: Interaction Effect with Analyst Forecast Dispersion

Panel (a) shows the average excess returns of a two way portfolio sort between O/S vs analyst forecast dispersion ($Disp$). Following Diether, Malloy, and Scherbina (2002), I define $Disp$ as the standard deviation of analyst forecasts divided by the absolute value of the mean forecast. The right most column shows the average excess return for each quintile of O/S . It is not necessarily the same as the simple average across each row because the quintiles of O/S are not independent of the quintiles of $Disp$. Panel (b) illustrates the same interaction effect using a Fama-MacBeth regression. Unreported controls are log market capitalization, log book to market ratio, the CAPM beta, returns over the past 12 months (excluding the most recent month), returns over the past month, and turnover, all cross-sectionally de-meanned. See Table 2.1 for full list of variable definitions. Data frequency is monthly, spanning 1996 to 2013, and has 223,049 firm-month observations, after merging with analyst forecast data.

(a) Two Way Portfolio Sort of Average Excess Returns						
	$Disp$ Low	2	3	4	$Disp$ High	All Stocks
(O/S) Low	0.89	0.84	1.09	0.97	0.94	0.95
2	0.80	0.57	0.35	0.39	0.53	0.54
3	0.57	0.52	0.40	0.14	0.28	0.39
4	0.64	0.28	0.25	-0.22	-0.20	0.14
(O/S) High	0.31	0.30	0.06	-0.03	-0.32	-0.02
(O/S) 5-1	-0.57**	-0.54**	-1.03***	-1.00***	-1.25***	-0.97***
t-statistic	(-2.3)	(-2.0)	(-3.2)	(-2.7)	(-3.4)	(-4.2)

(b) Fama-MacBeth Regression on Individual Stock Returns

$$R_{i,t+1}^e = b_{0,t} + \lambda_1 \cdot Qtile(O/S)_{i,t} + \lambda_2 \cdot Qtile(O/S)_{i,t} \times Qtile(Disp)_{i,t} + \lambda_3 \cdot Qtile(Disp)_{i,t} + b_1 \cdot X_{i,t} + \epsilon_{i,t+1}$$

	(1)	(2)	(3)
VARIABLES			
Qtile(O/S)	-0.15*** (-3.7)	-0.14*** (-3.9)	-0.14*** (-3.9)
Qtile(O/S) \times Qtile($Disp$)			-0.04*** (-3.2)
Qtile($Disp$)		-0.03 (-0.7)	-0.03 (-0.7)
Observations	223,049	223,049	223,049
R-squared	0.080	0.083	0.084
Number of months	211	211	211
Controls	Y	Y	Y

t-statistics in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

significant; however, I have verified that it has a negative statistically significant effect over a larger time span, confirming prior work using *Disp*. In Column (3), I add the interaction term $Qtile(O/S) \times Qtile(Disp)$. It is statistically significantly negative with $\lambda_2 = -0.04$ with $|t| > 3$. That is, higher O/S is associated with lower returns, but the relationship is even stronger in magnitude when there is higher analyst forecast dispersion.

In terms of the theory, this interaction effect appears to come from the fact that higher h is associated with more variation in the variables that drive O/S and expected returns. My model in Section 2.2 shows that variation in risk tolerance or in quantity of the underlying stock creates a negative correlation between O/S and expected returns. If the variation in the underlying variables is constant across quintiles of h , then the cross derivatives in Proposition 10 suggests that the return spread should fall with higher belief heterogeneity h . However, if there is more variation in the underlying variables when h is higher, then the return spread increases with h . The interaction effect shown in Table 2.7 suggests that the latter effect dominates. The latter effect also predicts that there should be more variation in O/S when h is higher. Table 2.2c suggests that is true, though the relationship is not perfectly monotonic.

Yet another possibility is that there are other dynamics between belief heterogeneity, O/S , and expected returns that my model in Section 2.2 does not capture. Nonetheless, the results in Table 2.7 suggest that there is a relationship between belief heterogeneity and the ability of O/S to forecast cross-sectional returns.

2.5 Conclusion

In this paper, I propose a disagreement model to understand the relative trading volume in an options market and its underlying market. Recent work by Johnson and So (2012) has found that the ratio of option trading volume to stock trading volume O/S negatively predicts stock returns in the cross section. They propose an asymmetric information model where investors with private negative information prefer to trade in the options market because they face shorting constraints in the underlying stock market. To prevent the

stock market in their model from reacting immediately, Johnson and So (2012) assume that investors in the underlying stock are somewhat irrational and do not react to options volume.

I show how a disagreement model in the spirit of Miller (1977) and Chen, Hong, and Stein (2002) also predicts a negative relationship between O/S and the expected return. I also document the following empirical findings and explain how the disagreement model is more consistent with these empirical results than an asymmetric information model: (1) O/S is positively correlated with analyst forecast dispersion, a proxy of belief heterogeneity, but is negative correlated with the bid-ask spread, a proxy of asymmetric information; (2) Both put option volume and call option volume are negative predictors of future returns in the cross section; (3) O/S is a negative predictor of cross-sectional returns over one year into the future; (4) The negative relationship between O/S and cross-sectional returns is stronger when analyst forecast dispersion is larger.

Chapter 3

Bond Fire Sales and Government Interventions

3.1 Introduction

During financial crises, financial institutions may be forced to sell assets at “fire sale” prices. As highlighted in Shleifer and Vishny (1992), these fire sales occur if the forced sales are correlated across different institutions. That is, if the forced sale were idiosyncratic, then one would expect another financial institution to step forward to purchase the asset at fair value. However, if the forced sale is industry-wide, then many financial institutions may be forced to dump assets at the same time, leading to a distressed price. Papers such as Krishnamurthy (2010), Merrill, et al. (2012), and Mitchell and Pulvino (2012) document empirical evidence of fire sales in the recent crisis.

Fire sales impose costs on the real economy by raising the hurdle rate on new investments. Instead of making new loans, banks may instead direct their capital towards purchasing fire sold assets (Shleifer and Vishny, 2010; He, Kang, and Krishnamurthy, 2010). This paper analyzes how the government should intervene to alleviate the consequences of fire sales. I build a model to examine two practical aspects: First, should the government directly purchase assets or offer leverage for the private sector to purchase assets? Second, should

the government intervene in the primary market of new issuances or in the secondary market of existing securities?

These questions are directly relevant to at least three large interventions by the U.S. government in the recent financial crisis. First, the Federal Reserve's Commercial Paper Funding Facility (CPFF) directly purchased commercial paper in the primary market. At its peak, the CPFF held \$350 billion of commercial paper (Adrian, Kimbrough, and Marchioni, 2011). Second, the original version of the U.S. Treasury's Troubled Asset Relief Program (TARP I) proposed purchasing assets in the secondary market. While the U.S. Treasury eventually shifted the TARP program to capital injections into the banks, the original proposal envisioned purchasing \$700 billion of "legacy" assets (CBO, 2009). Third, the Federal Reserve's Term Asset-Backed Securities Loan Facility (TALF) primarily provided leverage for the private sector to purchase newly-issued asset-backed securities. At its peak, TALF provided \$50 billion in loans and "the program was authorized to reach \$200 billion and at one point up to \$1 trillion in loan volume was envisioned" (Ashcraft, Malz, and Pozsar, 2012).

These questions are also relevant to the U.K. Funding for Lending Scheme (FLS), started in 2012 in response to the euro crisis. As of this writing, the program is still in effect with a total of GBP 55 billion in loans (BoE, 2014). Via the FLS, the U.K. government offers banks funding at below-market costs if they make new loans. While not directly structured as offering leverage for the purchase of newly issued bonds, the FLS program achieves that in essence because it conditions the amount of funding on the amount of new lending.

To some extent, my model is also relevant to the first round of quantitative easing (QE1) by the Federal Reserve in late 2008 to early 2010. In QE1, the Federal Reserve purchased \$1.25 trillion in mortgage-backed securities, \$300 billion Treasury securities, and \$172 billion in agency debt (Fischer, 2015). My model is closely related to the capital constraints channel of quantitative easing, where specialists with limited capital dominate the market for certain securities. However, it differs from other channels, such as the signaling channel, the safe asset channel, and the duration risk channel (Krishnamurthy and Vissing-Jorgensen, 2011

and 2013).

The intuition of my model is as follows. There are three types of agents: entrepreneurs, banks, and households. The entrepreneurs have investment opportunities, but lack capital. They raise the capital by selling bonds in the primary market. The households have capital, but I assume they cannot finance entrepreneurs directly. In this model, the households are passive and exist just to provide a frictionless benchmark. Finally, the banks have scarce capital to allocate between the primary market and secondary market. Here, “banks” represent the entire financial sector.

I generate a fire sale in the secondary market by assuming that the banks do not have sufficient net worth to purchase all the assets at fair value. As a result, the fire sale increases the hurdle rate on new bonds. The higher hurdle rate in turn imposes costs on the real economy because the level of investment in new projects is inefficiently low. This mechanism is similar to that of Shleifer and Vishny (2010), but my paper differs in its analysis of the government interventions. Whereas Shleifer and Vishny (2010) compare purchases in the secondary market vs equity injections, in contrast, this paper analyzes asset purchases in the primary and secondary market and also analyzes the effects of offering leverage to the private sector.

To alleviate the inefficiently low level of investment, the government raises funds by taxing the households and using those funds to intervene. The government’s objective is to minimize the hurdle rate on new bonds and raise the level of investment up to the frictionless benchmark, for a given dollar outlay. If the government chooses to directly purchase bonds, it does not matter if the government chooses to purchase securities in the primary vs secondary market. Intuitively, new bonds and existing bonds are perfect substitutes. This effect matches the casual intuition that the primary and secondary market are the same: one is the flow and the other is the stock. Hence, if the government intervenes more in one market, the private sector leans the opposite direction to offset the effect, so that on net only the total size of the intervention matters.

In contrast, my model shows that when offering leverage, the choice of primary vs

secondary market does matter. By focusing on offering leverage for the purchase of new securities, the government can lower the interest rate of new bonds and raise the level of investment. Intuitively, offering leverage breaks the perfect substitutability of the assets between the primary and the secondary market. Because banks have limited net worth, assets that can be pledged for more financing are more valuable than assets with the same cash flows, but less pledgeability. Hence, if the government offers leverage for purchasing bonds in the primary market, but not for purchasing bonds in the secondary market, then the assets in the two markets are no longer perfect substitutes and prices across the two markets deviate.

Furthermore, I show the conditions under which offering leverage leads to the same equilibrium prices as directly purchasing bonds, which allows me to compare the two interventions. I find that the government should focus on providing leverage to the primary market, over direct purchases in either market or providing leverage in the secondary market. My results show that offering leverage is not merely a safer way for the government to intervene, as there is no risk in my model. Rather, offering leverage has the additional benefit of allowing the government to more effectively target the intervention with less “leakage” into the secondary market.

While offering leverage allows the prices across the two markets to deviate, the prices do not deviate arbitrarily. The prices in the primary and secondary market are still linked because the bank equates the returns per unit of its equity capital across assets. Papers such as Brunnermeier and Pedersen (2009) and Garleanu and Pedersen (2011) also emphasize a similar linkage of prices across different markets when equity capital is scarce.

To illustrate, consider the following numerical example. Section 3.3.4 explains the details behind the calculations, but for now, I just describe the resulting bond yields. Suppose that in normal times, there is plentiful bank capital and new bonds have an interest rate of 0%. Then, a dislocation occurs in the secondary market and yields on old bonds rise. After banks arbitrage across the primary and secondary market, the interest rate in both markets rises to 30%, leading to an inefficient level of investment. To alleviate this problem,

the government can either directly purchase bonds or offer leverage for private agents to purchase bonds.

If the government chooses direct purchases, then its purchases lower the yields from 30% to 20%. The yields in the primary and secondary markets are the same, since banks arbitrage between the two markets. However, if the government instead chooses to offer leverage by reducing haircuts on new bonds to 50%, the yields in the primary market fall from 30% to 11% and the yields in the secondary market fall from 30% to 22%. Because newly issued bonds have twice the leverage of existing bonds, the bank earns the same leveraged return per unit of its capital, $11\% \times 2 = 22\%$ —the prices across the two markets are still linked in this way. Furthermore, since $11\% < 20\%$, offering leverage is more effective at lowering the rates on new loans than direct purchases.

My paper connects directly with Ashcraft, Garleanu, and Pedersen (2010) and Geanakoplos (2010), which both analyze how lowering haircuts raises asset prices. The main focus of both those papers differs from the focus of this paper though. Ashcraft, Garleanu, and Pedersen (2010) focuses on contrasting the policy tools of cutting interest rates vs lowering haircuts. Geanakoplos (2010) focuses on how to endogenously determine haircuts. Furthermore, both papers do not consider constraints on the size of the government intervention. Hence, Ashcraft, Garleanu, and Pedersen (2010) and Geanakoplos (2010) both conclude that offering leverage to both the primary and secondary market is better than just offering leverage to the primary market—even though the former requires a larger government intervention.

Section 3.2 describes the basic model and the equilibrium without government intervention. Section 3.3 describes modeling the government interventions along the two dimensions of direct purchases vs offering leverage and primary vs secondary market. It also contains a brief discussion of equity injections. Section 3.4 discusses potential extensions of the model and discusses how the model applies to U.S. and U.K. government policies implemented during the financial crisis and its aftermath. Section 3.5 concludes.

3.2 A Model of Fire Sales

In this section, I describe the setup of my model, establish the frictionless benchmark, and describe the constrained equilibrium without government intervention. My model has two time periods $t \in \{0, 1\}$, which I also refer to as “today” and “tomorrow.” There are three types of agents: entrepreneurs, banks, and households, each with a unit mass. There are three asset markets: a primary market for newly issued bonds with price p_x , a secondary market for previously issued bonds with price p_y , and a market for the risk-free asset. I normalize the gross risk-free rate to 1, by assuming it is elastically supplied at that rate.

Entrepreneurs have a project with diminishing marginal returns, where investing I dollars today yields $f(I)$ dollars tomorrow. Entrepreneurs have no net worth, so they raise financing by selling bonds today in the primary market. Each bond has face value 1 and sells at price p_x . I assume the projects are risk-less.

Entrepreneurs maximize investment returns

$$\max_I f(I) - \frac{I}{p_x}$$

Therefore, the equilibrium level of investment is I^* , as defined in the first order condition $f'(I^*) = \frac{1}{p_x}$. Intuitively, the entrepreneur equates the marginal return on investment with the marginal cost of capital. Since the entrepreneur raises I^* dollars, the supply of new bonds in the primary market is I^*/p_x .

Banks are active in all three asset markets: the market for the risk-free asset, the primary market for newly issued bonds, and the secondary market for previously issued bonds. Banks have limited net worth and must finance the purchases out of their initial net worth because they suffer from debt overhang. If there were no debt overhang, banks would issue equity to raise more capital. However, debt overhang causes debt is the marginal claimant in some states of the world. As a result, equity bears the full cost of equity issuances, but debt receives part of the benefits. This transfer of value is equivalent to a tax and dampens the bank’s willingness to issue equity (Myers, 1977). I focus on the case where the debt overhang constraint binds, i.e. the marginal return on capital is not high enough for the

equity holders to want to issue new equity. I assume that the banks' initial capital comes from a group of agents different from the households or the entrepreneurs.¹

Banks have an initial net worth of N^B at time 0 and aim to maximize their net worth at time 1. The bank purchases x^B units of bonds in the primary market at a unit price of p_x , y^B units of bonds in secondary market at a unit price of p_y , and z^B units of the risk-free asset at a unit price of 1. Therefore, banks maximize:

$$\max_{x^B, y^B, z^B} x^B + y^B + z^B \quad (3.1)$$

subject to the budget constraint:

$$N^B = p_x x^B + p_y y^B + z^B \quad (3.2)$$

To eliminate price feedback effects from changes in p_y onto the banks' net worth N^B , I assume that the existing bonds in the secondary market are initially held by an "old generation." This old generation dies at the end of $t = 0$ so they sell all the bonds in the secondary market to the banks, then consume the proceeds.

Households have a large endowment each period and maximize their consumption over the two time periods. They have linear utility $c_{t=0} + c_{t=1}$. In the frictionless benchmark, banks issue equity and households purchase it at fair value. However, households cannot directly participate in the primary or secondary market for bonds. Hence, when banks suffer from debt overhang, the households only hold the risk-free asset.

To establish a frictionless benchmark, I first consider the equilibrium without debt overhang. When the banks do not suffer from debt overhang, they issue equity to the households to raise capital. Since the households have a discount factor of 1 and since there is no risk, in the frictionless benchmark, $p_x^{*FB} = p_y^{*FB} = 1$ and the first best level of investment is

$$f'(I^{*FB}) = 1$$

¹In the long-run, new firms that do not suffer from debt overhang enter the market, but I assume that such new capital is slow moving and enters the market at a time scale beyond the model (Mitchell, Pedersen, and, Pulvino, 2007; Duffie, 2010).

Intuitively, in the frictionless benchmark, the gross cost of capital is 1 and the entrepreneurs invest until the marginal benefit of investment is 1.

3.2.1 Constrained Equilibrium with Debt Overhang

Let Q denote the outstanding stock of bonds in the secondary market. If the debt overhang constraint binds, then the banks do not have enough net worth to bring the price of the bonds in the primary market or the secondary market to the risk-free price, i.e. $p_x < 1$ and $p_y < 1$. Because the risk-free asset is elastically supplied at a price of 1, the banks do not hold any of the risk-free asset ($z^{*B} = 0$) when there is debt overhang. The following system of equations determines the equilibrium. First, there are the first order conditions of the bank and the entrepreneur:

$$\frac{1}{p_x} = \frac{1}{p_y} \quad (3.3)$$

$$f'(I) = \frac{1}{p_x} \quad (3.4)$$

Second, there are the market clearing conditions in the primary and secondary market:

$$\frac{I}{p_x} = x^B \quad (3.5)$$

$$Q = y^B \quad (3.6)$$

and the budget constraint $N^B = p_x x^B + p_y y^B$.

Since this equilibrium has no government intervention, I let I^{*NoGov}, p^{*NoGov} denote the equilibrium level of investment and the equilibrium bond price. Banks invest across both the primary market and the secondary market for bonds. Hence, at the interior solution, banks equate the return on capital across the two markets, $p_x = p_y = p$. To model a firesale, I assume that the banks have insufficient net worth to bring the price of the bonds to 1. More formally, under the parameter restriction $N^B < I + Q$, we get:

$$p_x^{*NoGov} = p_y^{*NoGov} < 1$$

Since $f'(I^{*NoGov}) > 1$, the level of investment is inefficient $I^{*NoGov} < I^{*FB}$.

For completeness, I also note a potential corner solution in this model. Suppose that the yields on new bonds cannot exceed a certain amount, or equivalently, there is a minimum price p_x^{min} such that $\frac{1}{p_x} < \frac{1}{p_x^{min}}$. In this situation, if the fire sale in the secondary market is sufficiently severe, the banks' optimal action is to use all of its capital to scoop up deeply discounted assets in the secondary market and allocate no capital to the primary market. In the language of the model, $\frac{1}{p_x^{min}} < \frac{1}{p_1}$, so Equation 3.3 fails to hold. In this corner solution, $I^* = 0$ and the price of the bonds in the secondary market is $p_y = N^B/Q$. At this corner solution, the level of investment is also too low since $0 < I^{*FB}$. Shleifer and Vishny (2010) consider a similar type of corner solution in their analysis of equity injections vs secondary market asset purchases.

There are at least two possible reasons why there may be a maximum possible yield on new bonds. First, we know that the entrepreneur's marginal return on investment cannot exceed $f'(0)$, the marginal return on the first dollar of new investment. Therefore, if $f'(0)$ is bounded and if the gross interest rate $\frac{1}{p_y}$ exceeds $f'(0)$, it is optimal for the entrepreneur to undertake zero projects. Second, one could also imagine a model with different types of entrepreneur quality and where banks have asymmetric information on the entrepreneurs' type. In this alternate setting, due to the asymmetric information, there may be credit rationing with a maximal interest rate (Stiglitz and Weiss, 1981).

In both the interior solution and the corner solution, investment is inefficiently low. Intuitively, due to limited bank capital, the fire sale imposes costs on the real economy because it raises the hurdle rate on new projects. When there is a fire sale, it is more attractive for banks to purchase the discounted bonds in the secondary market, instead of lending to entrepreneurs for new projects. The banks re-allocate capital from the primary market to the secondary market until the returns across the two markets reach the constrained equilibrium with I^{*NoGov}, p^{*NoGov} .

This model also illustrates another cost of universal banking. In this model, if the government barred the bank from participating in the secondary market, there would be more investment (though, admittedly, it would somewhat exacerbate the fire sale in the

secondary market). This model implies that, if banks are universal, then random arbitrage dislocations can draw away scarce capital from new loans. For example, suppose a bank primarily focuses on extending new commercial loans. If it can also trade in other markets, a dislocation in the on-the-run/off-the-run Treasury spread can cause the bank to instead use its capital to pursue this obscure arbitrage instead of making new commercial loans.

3.3 Government Interventions

This section analyzes the effects of government intervention on the constrained equilibrium. The government aims to bring the level of investment back towards the frictionless benchmark I^{FB} . The government raises $\tau_x + \tau_y$ dollars by taxing the households. The distortionary effects of taxation is one potential microfoundation for the government to constrain the size of its intervention. The government uses τ_x dollars to intervene in the primary market and τ_y dollars in the secondary market. I contrast two types of interventions: (1) intervening by directly purchasing bonds vs intervening by offering leverage/reducing haircuts on bonds; and (2) intervening in the primary market vs intervening in the secondary market.

3.3.1 Direct Purchases

I first consider the effect of the government directly purchasing of bonds. To model the direct purchase, suppose the government purchases τ_x dollars worth of bonds in the primary market and τ_y dollars worth of bonds in the secondary market, i.e. $\frac{\tau_x}{p_x}$ units of bonds and $\frac{\tau_y}{p_y}$ units of bonds, respectively. The first order conditions of the banks and the entrepreneurs is the same as before:

$$\frac{1}{p_x} = \frac{1}{p_y} \quad (3.7)$$

$$f'(I^*) = \frac{1}{p_x} \quad (3.8)$$

The market clearing conditions change to include the government's purchases:

$$\frac{I^*}{p_x} = x^B + \frac{\tau_x}{p_x} \quad (3.9)$$

$$Q = y^B + \frac{\tau_y}{p_y} \quad (3.10)$$

I assume that the government intervention is not so large that it fully offsets the fire sale. The parameter restriction $N^B + (\tau_x + \tau_y) \leq I^{*Direct} + Q$ implies $p^{*Direct} = p_x^{*Direct} = p_y^{*Direct} \leq 1$.

Proposition 11. *For a fixed total intervention size of $\tau_x + \tau_y$, the equilibrium price $p^{*Direct}$ is the same, regardless of the government's choice of primary vs secondary market (τ_x vs τ_y). That is, when the government engages in direct purchases, only the total size of the intervention $\tau_x + \tau_y$ matters, not the breakdown.*

Fundamentally, this result comes from the fact that banks arbitrage between the primary and secondary market. The primary market is the flow and the secondary market is the stock. Since they are the same, only the total size of the intervention $\tau_x + \tau_y$ matters, not the breakdown. If the government intervenes more in the primary market, then the banks offset their actions by purchasing more assets in the secondary market. Since the return from direct purchases is $\frac{1}{p^{*Direct}} > 1$, this intervention is revenue positive, which allows the government to pay back the funds it borrowed from the households.

3.3.2 Equity Injections

In this model, an equity injection is the same as a direct purchase. Note that the government would likely have to force the equity injection. Because of the debt overhang, the banks are unwilling to issue equity at market terms. Hence, the government would either have to force the banks to take capital anyway or offer above market terms. Suppose the government forces the banks to accept an equity injection of $\tau_x + \tau_y$. This injection increases the banks' net worth from N^B to $N^B + \tau_x + \tau_y$.

Proposition 12. *An equity injection of $\tau_x + \tau_y$ leads to the same equilibrium price as a direct purchase of $\tau_x + \tau_y$, i.e. $p^{*EqtyInjection} = p^{*Direct}$*

Intuitively, if the government injects equity to the banks, the banks will still arbitrage between the primary and secondary markets.

3.3.3 Offering Leverage for Purchases

In direct purchases, the choice of intervening in the primary vs secondary market does not matter because a newly issued bond is a perfect substitute for a previously issued bond. Here, I show how the government can break the perfect substitutability between the primary and secondary market by offering leverage for bond purchases. Here is the intuition: Suppose the government only accepts primary market bonds as collateral for the leverage it offers. Because the banks suffer from debt overhang, the ability to pledge bonds as collateral for leverage is valuable. Hence, the primary and secondary market bonds are no longer identical and prices across the two markets can differ. Total government outlays will again be $\tau_x + \tau_y$, but the breakdown will now affect the equilibrium prices.

First, I describe how the government offers leverage. Recall that one unit of a newly-issued bond costs p_x . Until now, banks had to pay for p_x entirely out of their net worth.² Suppose the government now allows banks to pledge a bond as collateral for a loan. Let h_x denote the percentage haircut on this loan, so that each bank pays for $h_x p_x$ out of its own net worth and borrows the rest $p_x - h_x p_x$ from the government. Similarly, the government allows the banks to pledge the bonds from the secondary market with haircut h_y . Since the projects are risk-less, I assume that the government offers leverage to banks at a 0% net interest rate.

As before, let τ_x denote the dollars the government spends in the primary market and τ_y denote the dollars the government spends in the secondary market. For each unit of bond that the bank purchases in the primary market, the government provides a collateralized loan for $p_x - h_x p_x$ dollars, and similarly for the secondary market. Also, as before, let x_i^B denote the number of units of bond the bank purchases in market i . Therefore, the

²Implicitly, I assume that private sector haircuts are exogenously set at 100%. It is straightforward to adjust my model to account for a different exogenous private sector haircut. However, making the haircuts endogenous requires a more complex framework such as Geanakoplos (2010).

government outlays $\tau_x = (p_x - h_x p_x) \cdot x^B$ in the primary market and $\tau_y = (p_y - h_y p_y) \cdot y^B$ in the secondary market.

Without the government loan, the gross return to purchasing the bond in market i was $\frac{1}{p_i}$. With the government loan, the gross return is now $\frac{1-(p_i-h_i p_i)}{h_i p_i}$. The banks' maximization problem is:

$$\max_{x^B, y^B, z^B} [1 - (p_x - h_x p_x)] \cdot x^B + [1 - (p_y - h_y p_y)] \cdot y^B + z^B + \lambda (N^B - (h_x p_x) x^B - (h_y p_y) y^B - z^B)$$

where x^B denotes the number of units of primary market bonds, y^B denotes the number of units of secondary market bonds, and z^B denotes the number of units of the risk-free bond. At the interior solution, the first order condition is $\frac{1-(p_x-h_x p_x)}{h_x p_x} = \lambda = \frac{1-(p_y-h_y p_y)}{h_y p_y}$. That is, the bank again equates the return on its capital across the two markets.

If the government offers different leverage terms in the primary market and secondary market ($h_x \neq h_y$), then the prices across the two markets differ ($p_x \neq p_y$). For example, suppose the net return on purchasing a bond in the secondary market is 22% (i.e. gross return is $\frac{1}{p_y} = 122\%$). Suppose also that the government does not offer any leverage against secondary market bonds ($h_y = 1$), but allows banks to borrow against half the value of the bonds in the primary market ($h_x = \frac{1}{2}$). Then, the unleveraged return in the primary market is 11% (i.e. $\frac{1}{p_x} = 111\%$). The banks' first order condition requires that the return on the banks' leveraged capital be equal across markets, but the unleveraged return can differ due to the leverage offered by the government.

For a given set of exogenous variables $\{N^B, Q, \tau_x + \tau_y\}$, the government's choice of h_x pins down $I^{*Lev}, p_x^{*Lev}, p_y^{*Lev}, h_y$ via the following system of equations:

$$\frac{I^*}{p_x} = x^B \quad (3.11)$$

$$Q = y^B \quad (3.12)$$

$$\frac{1 - (p_x - h_x p_x)}{h_x p_x} = \frac{1 - (p_y - h_y p_y)}{h_y p_y} \quad (3.13)$$

$$\tau_x + \tau_y = (p_x - h_x p_x) \cdot \frac{I^*}{p_x} + (p_y - h_y p_y) \cdot Q \quad (3.14)$$

$$f'(I^*) = \frac{1}{p_x} \quad (3.15)$$

Since I assume that the government intervention is not enough to fully offset the fire sale ($p_i \leq 1$), we have that $N^B + (\tau_x + \tau_y) < I^{*Lev} + Q$. Also, since no haircut cannot exceed 1, we also have the parameter restriction that $h_y = \frac{(1-h_x)I^* + p_y Q - (\tau_x + \tau_y)}{p_y Q} < 1$, which is equivalent to $(1 - h_x)I^* < \tau_x + \tau_y$.

Proposition 13 describes how offering leverage differs from direct purchases.

Proposition 13. *For a fixed total government outlay $\tau_x + \tau_y$, the government's choice of h_x affects the equilibrium price $\frac{\partial p_x^{*Lev}}{\partial h_x} < 0$. That is, when the government offers leverage for bond purchases, the breakdown of τ_x, τ_y does affect the equilibrium price p_x^* , in contrast to direct purchases by the government.*

Intuitively, when the government intervenes with direct purchases, the bonds in the primary and secondary markets assets continue to be perfect substitutes. For this reason, only the total size of the direct purchases matters and not the breakdown. In contrast, when the government offers leverage, it can break the perfect substitutability. Because banks have limited capital, the ability to pledge an asset as collateral has value. Two assets with the same cash flows, but different pledgeability will have different prices (see also Garleanu and Pedersen, 2009). By offering different haircuts across the two markets, the government can more tightly focus the intervention in the primary market, with less “leakage” into the secondary market.

This model also shows that offering leverage is not merely a safer way for the government to intervene. In a model with risk, the government may be concerned about potential losses. For the same dollar outlay, direct purchases expose the government to more risk than just offering leverage to the private sector. Since the government rarely intervenes with the goal of making money from scooping up distressed assets, the government may choose to accept a lower return by offering leverage in exchange for lower risk. While that is a legitimate concern, this model illustrates that offering leverage has other effects, as there is no risk in this model.

Proposition 14 explains the special case when offering leverage is equivalent to direct purchases. If we combine Proposition 13 and Proposition 14, we can conclude that when

$h_x > h_y$, offering leverage more effectively lowers the hurdle rate in the primary market than direct purchases.

Proposition 14. *Given the fixed total outlay $\tau_x + \tau_y$, the government's choice of haircut in the primary market h_x pins down the haircut in the secondary market h_y . If the government chooses a value of h_x such that $h_x = h_y$, then $p_x^{*Lev} = p_y^{*Lev}$. Furthermore, let $p^{*Direct}$ denote the equilibrium price under direct purchases. When leverage in the primary and secondary markets is the same ($h_x = h_y$), offering leverage is the same as direct purchases ($p_x^{*Lev} = p_y^{*Lev} = p^{*Direct}$).*

3.3.4 Numerical Simulation: Offering Leverage vs Direct Purchases

In this subsection, I use a numerical simulation to illustrate the propositions above. In this numerical simulation, I assume that the entrepreneur's investment function has the form $f(I) = \ln I$, which implies that $I = p_x$. I set the exogenous parameters as $N^B = 10, Q = 12, \tau_x + \tau_y = 0.75$ (i.e. to make it realistic, suppose units are trillions of dollars). The numerical example in Section 3.1 comes from this simulation.

In the frictionless benchmark, there is no debt overhang so $p^{*FB} = p_x^{*FB} = p_y^{*FB}$ and:

$$p^{*FB} = 1 \quad (3.16)$$

If there is debt overhang, but no government intervention, then $p^{*NoGov} = p_x^{*NoGov} = p_y^{*NoGov}$ and:

$$p^{*NoGov} = \frac{N^B}{1 + Q} \quad (3.17)$$

For the given parameters, $p^{*NoGov} = 0.76$. This price corresponds to the 30% yield mentioned in the numerical example in Section 3.1.

Suppose the government intervenes by purchasing $\tau_x + \tau_y$ dollars of bonds. As shown in Proposition 11, $p^{*Direct} = p_x^{*Direct} = p_y^{*Direct}$. Given the functional form of the entrepreneur's investment function, we have:

$$p^{*Direct} = \frac{N^B + (\tau_x + \tau_y)}{1 + Q} \quad (3.18)$$

For the given parameters, $p^{*Direct} = 0.83$ for all values of h_x . This price corresponds to the 20% yield mentioned in the numerical example in Section 3.1.

Now suppose the government intervenes by offering leverage. Because the total size of the intervention is fixed at $\tau_x + \tau_y$, the government's choice of haircut in the primary market h_x pins down the other outcome variables $p_x^{*Lev}, p_y^{*Lev}, h_y$. Given the functional form,

$$p_x^{*Lev} = \frac{N^B}{N^B - h_x(N^B + \tau_x + \tau_y - Q - 1)} \quad (3.19)$$

$$p_y^{*Lev} = \frac{N^B(N^B + \tau_x + \tau_y - 1) - h_x(N^B + \tau_x + \tau_y)(N^B + \tau_x + \tau_y - Q - 1)}{Q(N^B - h_x(N^B + \tau_x + \tau_y - Q - 1))} \quad (3.20)$$

$$h_y = \frac{N^B(h_x(N^B + \tau_x + \tau_y - Q) - N^B)}{N^B + (N^B + \tau_x + \tau_y)(h_x(N^B + \tau_x + \tau_y - Q - 1) - N^B)} \quad (3.21)$$

In the numerical example in Section 3.1, $h_x = 0.5$, which implies that $p_x^{*Lev} = 0.90$ (or 11% yield) and $p_y^{*Lev} = 0.82$ (or 22% yield).

Figure 3.1 plots the outcome variables from offering leverage $\{p_x^{*Lev}, p_y^{*Lev}, h_y\}$ as a function of h_x . Lower haircuts are equivalent to more leverage. The haircut for the secondary market asset cannot exceed 1 so the horizontal axis does not start at 0. For comparison, I also plot the equilibrium price of the direct purchases intervention $p^{*Direct}$. To avoid over-cluttering the figure, I do not plot the equilibrium price of the frictionless benchmark p^{*FB} or the equilibrium price without government intervention p^{*NoGov} , but they are important points of comparison as well.

In general, we observe that $p_x^{*Lev} \neq p_y^{*Lev}$. Intuitively, by offering different haircuts, the government breaks the perfect substitution across the primary market and the secondary market, which allows the prices to deviate. The sole exception is that $p_x^{*Lev} = p_y^{*Lev} = p^{*Direct}$ (the collinear point in Figure 3.1) when $h_x = h_y$. This collinear outcome is a result of Proposition 14, which states that implementing equal haircuts for all bonds is equivalent to direct purchases.

The figure also illustrates the results of Proposition 13. First, lowering haircuts (increasing leverage) on bonds in the primary market increases p_x^{*Lev} . Since $I^* = p_x$, this intervention bring the economy closer to the frictionless benchmark level of investment. In this simulation,

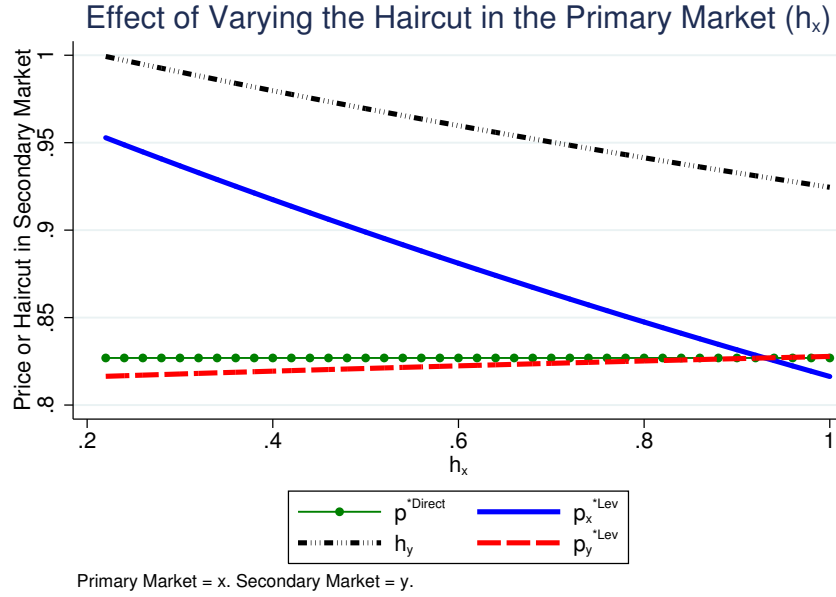


Figure 3.1: Numerical Simulation of Direct Purchases vs Offering Leverage

This graph uses a numerical simulation to contrast the effects of a government intervention offering leverage on bonds vs a government directly purchases assets. The graph depicts how different variables in my model change as the government changes h_x , the haircut it offers for the bond in the primary market. Lower haircuts are equal to more leverage. The government's total dollar outlay is constrained to $\tau_x + \tau_y$. The outcome variables for the offering leverage intervention are $\{p_x^{*Lev}, p_y^{*Lev}, h_y\}$, where p_x^{*Lev} is the price of the bond in the primary market; p_y^{*Lev} is the price of the bond in the secondary market; and h_y is the haircut in the secondary market. The outcome variable for the direct purchases intervention is $p^{*Direct}$, which is the price of the bond in the primary and secondary market when the government instead purchases $\tau_x + \tau_y$ dollars worth of assets. The horizontal axis does not start at 0 because the haircut for the asset in the secondary market cannot exceed 1. The parameters for simulation are $N^B = 10$, $Q = 12$, $\tau_x + \tau_y = 0.75$. For comparison, in the (unplotted) frictionless benchmark, $p^{*FB} = p_x^{*FB} = p_y^{*FB} = 1$. Also for comparison, if there is debt overhang, but no government intervention, $p^{*NoGov} = p_x^{*NoGov} = p_y^{*NoGov} = 0.76$ (also unplotted).

the government maximizes the level of investment by concentrating its efforts on maximize leverage in the primary market, instead of spreading the leverage out evenly. In particular, when the government sets $h_x = 0.2$, it achieves $p_x^{*Lev} = 0.96$.

Even if the government maximizes the leverage available to the primary market asset, because haircuts h_y on the secondary market asset cannot exceed 1, prices and investment in the primary market still fall short of the first-best level, i.e. $p_x^{*Lev} < p^{*FB} = 1$ and $I^{*Lev} < I^{*FB} = 1$.

Finally, in general, $p_x^{*Lev} > p^{*Direct} > p_y^{*Lev}$. That is, increasing p_x^{*Lev} slightly increases the size of the fire sale in the secondary market relative to direct purchases. However, relative to the no intervention case, the secondary market is always better off, i.e. $p_y^{*Lev} \geq p^{*NoGov}$ for all values of h_x .

3.4 Discussion

3.4.1 Potential Extensions of the Model

Risk and Asymmetric Information: In this model, the projects and the bonds are risk-less. Allowing for risk changes the model in at least two ways. First, the model would focus on excess returns instead of total returns. In my current model, there is no risk, so doubling the leverage increases the return, but does not increase the risk (e.g. compare the bank's return $\frac{1-(p_x-h_x p_x)}{h_x p_x}$ for $h_x = 1.0$ and $h_x = 0.5$). Second, because the bank suffers from debt overhang, if there is risk in the model, the bank has the temptation to risk-shift at the expense of its creditors (Jensen and Meckling, 1976). Furthermore, if the government knows less about the assets than the bank, banks are incentivized to sell the bad assets to the government. In such a situation, offering leverage may be more attractive than direct purchases because a collateralized loan is less informationally sensitive than directly purchasing the asset.

Price Feedback Effects: My model did not allow for price feedback effects, by assuming that the existing bonds were initially held by an "old generation." Restoring price feedback effects may generate other interesting results. In particular, it may change the government's

objective function. In the current model, the government aims to restore investment to the frictionless level and does not directly care about the fire sale in the secondary market. As we saw in Figure 3.1, both offering leverage and direct purchases increase investment, relative to the no intervention case (i.e. $p_x^{*Lev}, p_y^{*Lev}, p^{*Direct} > p^{*NoGov}$). Moreover, focusing leverage in the primary market more effectively stimulates new lending than direct purchases $p_x^{*Lev} > p^{*Direct}$. However, relative to direct purchases, focusing leverage in the primary market somewhat exacerbates the secondary market fire sale (i.e. $p_x^{*Lev} > p^{*Direct} > p_y^{*Lev} > p^{*NoGov}$). If there are price feedback effects, then the government may place more weight on the price in the secondary market because it affects the net worth of banks in the system. The weight the government places on the secondary market in turn affects the type of intervention the government chooses.

Dynamic Considerations: My model is a two period model. However, one could also consider a multi-period model. For example, suppose there is a pre-period $t = -1$ where banks make loans before hitting the fire sale at $t = 0$. The potential for future fire sales can raise the hurdle rate in the pre-period of $t = -1$, as banks attempt to hold “dry powder” to seize future deals. The Shleifer and Vishny (1997) model of limited arbitrage features a similar “dry powder” effect. By committing to intervene if a fire sale occurs, the government can reduce the incentive for banks to hold such “dry powder.” In my model, the government makes money on the direct purchases and breaks even on offering leverage, so it does not face dynamic consistency issues related to the commitment to intervene, in general.

In a dynamic model, the price feedback effects I discussed above can create a dynamic consistency issue because the prices at $t = 0$ affect the net worth at $t = 0$. The model in this paper only has two periods, so the government focuses on the primary market. Taken literally, the government should focus all of its intervention on offering leverage to new bonds. But, in a dynamic model, the bonds in the primary market at $t = -1$ become the bonds in the secondary market at $t = 0$. Furthermore, in a dynamic model, the units of bonds that the bank purchases at $t = -1$ affects their net worth at $t = 0$. So, on one hand,

at $t = -1$, to encourage investment in that period, the government would like to promise that it will support prices in the secondary market if there is a firesale at $t = 0$. On the other hand, once it is at $t = 0$, the government no longer cares about the sunk investment at $t = -1$ and would like to focus all of its intervention on the primary market. Hence, if the banks believe the government is not credible in its commitment to support the secondary market at $t = 0$, then that will increase the interest rate at $t = -1$.

If the government does not insist that it make money on its interventions (e.g. perhaps the social benefit of raising the level of investment exceeds the loss on the intervention), then the dynamic model interacts with the riskiness of the loans highlighted above. For example, if the banks anticipate that the government will intervene and over-pay for assets at $t = 0$, then they may make loans to projects of worse quality in the pre-period $t = -1$. This degrades the overall quality of the bonds at $t = 0$, since last period's new issuances become this period's existing bonds.

Allowing Entrepreneurs to Participate in the Bond Market: In my model, I assumed that only the banks participate in the bond market. One could also consider an extension where entrepreneurs can participate in the bond market. In this situation, entrepreneurs have the incentive to sell a bond in the primary market and then use the proceeds to purchase another bond and pledge it as collateral to the government for even more financing. Note however that the entrepreneur is indifferent between the primary and the secondary market because the two markets are in equilibrium. The unleveraged returns in the two markets differ, but the leveraged returns are the same.

Endogenous Haircuts: In this model, the government can choose any h_x as long as $0 < \{h_x, h_y\} \leq 1$. However, the endogenous level of haircuts from the private sector may influence the government's choice. For example, suppose that specialized creditors typically offer leverage for these bond purchases. Then, if these specialized creditors simply do not have enough capital during the fire sale, then the government is actually alleviating a fundamental friction. In contrast, suppose private sector haircuts are actually at the right level to deter moral hazard. In this case, because the government most likely cannot monitor

borrowers as well as the private sector, government intervention will not help. A more complete model of haircuts might also imply alternative policies. For example, if the true friction is that specialized creditors do not have enough capital, perhaps the government should recapitalize these specialized creditors instead of directly offering leverage to specific bonds.

3.4.2 U.S. and U.K. Government Interventions

Here, I discuss how my model relates to actual government interventions during the recent financial crisis and its aftermath. I discuss one program by the U.S. Treasury, three programs by the Federal Reserve, and one program by the Bank of England in partnership with the U.K. Treasury. I also discuss how potential considerations outside my model may have shaped the design of those programs.

U.S. Treasury's Troubled Asset Relief Program (TARP I) and Federal Reserve's Commercial Paper Funding Facility (CPFF): During the 2008-2009 financial crisis, the U.S. government pursued two types of direct purchase programs. One was considered, but not ultimately implemented and the other was implemented. In the mortgage security market, the original proposal of the Trouble Asset Relief Program (TARP I) in October 2008 aimed to purchase "residential or commercial mortgages and any securities, obligations, or other instruments that are based on or related to such mortgages" (CBO, 2009). Moreover, the program stated that it aimed to purchase securities in the secondary market, specifically assets issued before March 14, 2008. In the end, the U.S. Treasury ended up pivoting this program into capital injections into the major banks, sometimes called TARP II, and so the U.S. Treasury did not ultimately purchase mortgage securities. In the commercial paper market, the Federal Reserve created the CPFF in October 2008. In contrast to TARP I, this program focused on newly issued commercial paper, not already issued securities. At its peak, the CPFF held \$350 billion of commercial paper (Adrian, Kimbrough, and Marchioni, 2011).

My model suggests that the choice of primary vs secondary market did not matter

for either of these programs if the goal was to stimulate the issuance of new bonds in the primary market. Proposition 11 discusses how only the total size of the intervention across the two markets combined matters. Intuitively, if the government intervenes more in one market, the banks lean the opposite direction to equilibrate the returns across the two markets. However, in the actual history, policy makers may have weighed other considerations not in my model. For examples, by directly purchasing commercial paper, the Federal Reserve's CPFF may be able to stem Diamond and Dybvig (1983) type run-risk, even though it may be less effective than offering leverage at lowering the interest rates on new loans.

Adrian, Kimbrough, and Marchioni (2011) provide empirical evidence on the Federal Reserve's CPFF. Their empirics unfortunately do not have an identification strategy to separate between correlation and causation, but it does provide some background facts. They document that yields on CPFF-eligible securities declined by about 150bp from October 2008 to December 2009. They further argue that the CPFF was successful in reviving the market for longer-term commercial paper. After Lehman's bankruptcy in September 2008, money market funds shortened the duration of their assets and the issuance of commercial paper with maturities exceeding one week "collapsed." Issuance in this market did not recover until the start of the CPFF.

However, Adrian, Kimbrough, and Marchioni (2011) also document that yields on lower-rated securities not eligible for CPFF rose by 20bp from October 2008 to December 2009. This evidence does not speak directly to my model because it does not study the differential pricing in the primary and secondary markets. However, the evidence does suggest that the CPFF lowered yields, but primarily for eligible securities. This suggests there is meaningful market segmentation in the commercial paper market as the banks did not arbitrage between the higher-rated securities and the lower-rated securities. If the primary and secondary market is similarly segmented, then the Proposition 11 may not apply.

Federal Reserve's Term Asset-Backed Securities Loan Facility (TALF): In November 2008, the Federal Reserve also implemented TALF, which offered leverage for the purchase of

bonds. While several different types of collateral were eligible, the TALF program primarily focused on offering leverage for newly issued asset-backed securities. At its peak, TALF loaned out \$50 billion worth of financing and was prepared to loan out \$200 billion of financing. From the perspective of my model, the TALF program was more effective than TARP I or the CPFF at lowering the hurdle rate on new investments and increasing new investment. As shown in Proposition 13 and Proposition 14, offering leverage allows the government to break the perfect substitution between the primary and secondary market and focus the intervention on lowering the hurdle rate for new investments.

Ashcraft, Garleanu, and Pedersen (2010) provide empirical evidence on TALF. They study how the market prices of AAA-rated super senior bonds changed depending on its eligibility for the TALF program. The authors argue that the Federal Reserve's rejection did not convey information about the cash flows of the bond because the Federal Reserve only used bond data that was available to other investors. They find that bond yields jumped on the announcement of rejection and fell on the announcement of acceptance into the TALF program. These changes in yields are consistent with the effects of leverage in my model. Because bank capital is scarce, assets that are leveragable are more valuable and hence have lower yields. Ashcraft, Malz, and Pozsar (2012) argue that TALF revived the primary market for asset-backed securities. They highlight that issuance plummeted greatly in late 2008 and its reopening in early 2009 was dominated by TALF-eligible securities.

While TALF focused primarily on newly issued asset-backed securities, it also provided leverage for previously issued commercial mortgage backed securities. My model disagrees with the rationale for the secondary market aspect of TALF. In Ashcraft, Malz, and Pozsar (2012), researchers at the Federal Reserve Bank of New York argued that the secondary market program helped the primary market: "Secondary-market spreads constitute hurdle rates for new issuance, since potential investors have the choice of buying bonds in the secondary market rather than the new-issue market." While it is true that investors allocate capital across the two markets, Proposition 13 shows why that logic is incomplete. Investors seek to equilibrate return on *leveraged capital* across the two markets. Hence, by offering

leverage, the government can allow the return on *unleveraged capital* to differ across the two markets. For example, using the numerical simulation from Section 3.3.4, if the government offers 2x leverage in the primary market, then it is an equilibrium to have 11% yield in the primary market and 22% yield in the secondary market. As a result, it is more effective to focus leverage on newly issued assets, not legacy assets.

Bank of England and U.K. Treasury's Funding for Lending Scheme (FLS): In 2012, the Bank of England and the U.K. Treasury launched the Funding for Lending Scheme. As of this writing, the program is still in effect with a total of GBP 55 billion in loans (BoE, 2014). Whereas the Federal Reserve and U.S. Treasury programs were in response to the 2008-2009 financial crisis, the FLS was in response to the euro crisis. The FLS aims to increase bank lending by providing banks with funding at below market costs. Roughly speaking, the FLS allows a bank to initially borrow against 5% of its stock of loans in June 2012. Then, the FLS allows banks to borrow additional funds proportional to their net lending after June 2012.³ The stated goal of the FLS is to spur new lending in the United Kingdom by reducing the financing costs of U.K. banks.

My model supports the basic idea behind the FLS. The FLS incentivizes banks to extend new loans by conditioning the amount of cheap funding they can receive from the government on the amount of new loans they make. While structured somewhat differently, the FLS is analogous to the offering leverage on bonds in the primary market in my model. One small way the FLS differs from my model is that the scheme conditions the loans on the *net* amount of lending relative to June 2012, whereas my model allows banks to borrow against new lending in general. The FLS may be structured this way because of implementation concerns. My model focuses on liquid bonds, which are easier to offer leverage against on a per bond basis. However, the FLS focuses on lending more generally, including small business loans, which may be harder to offer leverage against on a per loan basis. As a result, the FLS program is more permissive in the type of collateral it accepts.

³This description necessarily abstracts from some of the implementation details. See the Bank of England's publication Churm, et al. (2012) for further details.

Instead of pledging the new bond itself as collateral, under the FLS, a bank can pledge any asset it has. Because of this flexible pledgeability, the FLS incentivizes the banks by conditioning the funding on the net lending.

However, my model disagrees with the way the FLS allows banks to borrow against their existing stock of loans. Rather, my model argues that they should just focus on lending against new loans. As discussed in Proposition 13, offering leverage is most effective in raising the level of investment when it is targeted at newly issued bonds. By targeting the newly issued bonds, the government gives these bonds extra collateral value, which allows them to focus their intervention in the primary market. Because the FLS allows banks to borrow against their existing stock of loans, a bank can borrow and use the additional funding to purchase more fire sold assets in the secondary market, which does not help new lending. In a Wall Street Journal opinion piece, Blinder (2012) more critically calls this aspect of the FLS a “straight giveaway to banks.”

Federal Reserve’s Large Scale Asset Purchases (LSAP): From 2008 to 2014, the Federal Reserve pursued multiple rounds of Large Scale Asset Purchases. In dollar amounts, the size of the LSAP programs vastly exceeded the other interventions, with \$3.6 trillion of securities purchases (Fischer, 2015). The LSAP programs consisted of three rounds of quantitative easing (which involved purchasing different combinations of Treasuries, agency debt securities, and mortgage-backed securities) and one maturity extension program (swapping short-term Treasuries for long-term Treasuries).

My model is closely related to the “capital constraints” channel of the LSAP programs (Curdia and Woodford, 2010; He and Krishnamurthy, 2013). In this channel, specialists with limited capital dominate certain markets, like the mortgage-backed securities market. Because the specialists have limited capital, purchases of the specialized securities by the central bank can lower yields. The direct purchases in my model have a similar effect. The banks have limited capital and the households do not participate in the bond market. For this reason, prices fall below the frictionless benchmark and the government can raise prices through direct purchases. Krishnamurthy and Vissing-Jorgensen (2011) argue that the

first round of quantitative easing, which involved primarily purchasing mortgage-backed securities, lowered yields by roughly 100bp through this channel.

However, my model does not speak to the other channels related to the LSAP programs, especially the ones related to the purchases of Treasuries. As surveyed by Krishnamurthy and Vissing-Jorgensen (2011, 2013), other channels include: the signaling channel (central bank purchases signal commitment to keep rates low); the safe asset channel (central bank purchases constrict the supply of safe assets, thereby lowering yields); duration risk channel (central bank purchases reduce the duration risk that the private sector must bear, thereby lowering yields), among others. In these channels, any agent can purchase the bonds—some agents even have a special preference for certain bonds (see the duration risk channel and the safe asset channel). In contrast, in the capital constraints channel, only the specialists purchase certain bonds. For this reason, the capital constraints channel is not likely to apply to the purchase of Treasuries. Empirical work has found that the Treasury-based LSAP programs also reduced the yields on Treasuries, though less than the amount they reduced the yields on mortgage-backed securities (Fischer, 2015).

3.5 Conclusion

This paper analyzes how governments can intervene to alleviate the real costs of fire sales. In this model, fire sales are costly because they raise the hurdle rate on loans for new projects. The model studies interventions that differ along two dimensions: (1) direct purchases of assets vs offering leverage for private agents to purchase assets and (2) intervening in the primary market for new issuances vs secondary market for existing assets. This paper differs from the previous literature by constraining the size of government intervention.

When the government intervenes using direct purchases, its choice of primary market purchases vs secondary market purchases does not matter since the assets in the two markets are perfect substitutes. While offering leverage may simply seem like a safer way to intervene, this model illustrates the unique aspects of offering leverage for bond purchases, beyond risk considerations. By offering security-specific leverage, the government breaks

the perfect substitutability between the primary and secondary markets. This ability allows the government to enhance the effectiveness of its interventions at stimulating new lending because less of the intervention “leaks” into the secondary market.

Bibliography

- [1] Adrian, T., Kimbrough, K., & Marchioni, D. (2011). The Federal Reserve's Commercial Paper Funding Facility. Federal Reserve Bank of New York Economic Policy Review, 25-39.
- [2] Andreou, P. C., Kagkadis, A., Maio, P. F., & Philip, D. (2014). Dispersion in Options Traders' Expectations and Stock Return Predictability. Working Paper.
- [3] Angrist, J. D., & Pischke, J. S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton University Press.
- [4] Ashcraft, A., Malz, A., & Pozsar, Z. (2012). The Federal Reserve's Term Asset-Backed Securities Loan Facility. Federal Reserve Bank of New York Economic Policy Review, 1-38.
- [5] Ashcraft, A., Garleanu, N., & Pedersen, L. H. (2011). Two monetary tools: Interest rates and haircuts. NBER Macroeconomics Annual 2010, 25, 143-180.
- [6] Avramov, D., Chordia, T., & Goyal, A. (2006). Liquidity and autocorrelations in individual stock returns. The Journal of Finance, 61(5), 2365-2394.
- [7] Bakshi, G., & Kapadia, N. (2003). Volatility risk premiums embedded in individual equity options: Some new insights. The Journal of Derivatives, 11(1), 45-54.
- [8] Banerjee, S., & Graveline, J. J. (2014). Trading in derivatives when the underlying is scarce. Journal of Financial Economics, 111(3), 589-608.
- [9] Bank of England (BoE). (2014). Funding for Lending Scheme extension: Usage and lending data. Accessed from website <http://www.bankofengland.co.uk/markets/Pages/FLS/extensiondata.aspx> during April 2015.
- [10] Bank for International Settlements. (2014). Statistical annex of September 2014 report. BIS Quarterly Review.
- [11] Black, F. (1975). Fact and fantasy in the use of options. Financial Analysts Journal, 31(4), 36-41.
- [12] Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. The Journal of Political Economy, 81(3), 637-654.

- [13] Ben-David, I., Franzoni, F., & Moussawi, R. (2014). Do ETFs increase volatility? NBER Working Paper No. w20071.
- [14] Blinder, A. (2012 July 31). An intriguing idea to encourage bank lending. Wall Street Journal Op-Ed.
- [15] Bollen, N. P., & Whaley, R. E. (2004). Does net buying pressure affect the shape of implied volatility functions?. *The Journal of Finance*, 59(2), 711-753.
- [16] Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430), 443-450.
- [17] Boyle, P. P., & Vorst, T. (1992). Option replication in discrete time with transaction costs. *The Journal of Finance*, 47(1), 271-293.
- [18] Brady, N. F., & United States Presidential Task Force on Market Mechanisms. (1988). Report of the presidential task force on market mechanisms. U.S. Government Printing Office.
- [19] Brown-Hruska, Sharon. (2006 March 28). Does the tail wag the dog? The case of natural gas markets. U.S. Commodity Futures Trading Commission, speech before the Natural Gas Roundtable of Washington.
- [20] Brunnermeier, M. K., & Pedersen, L. H. (2009). Market liquidity and funding liquidity. *Review of Financial studies*, 22(6), 2201-2238.
- [21] Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1), 57-82.
- [22] Campbell, J. Y., Grossman, S. J., & Wang, J. (1992). Trading volume and serial correlation in stock returns. *Quarterly Journal of Economics*, 108 (4), 905-939.
- [23] Campbell, J. Y., Lettau, M., Malkiel, B. G., & Xu, Y. (2001). Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. *The Journal of Finance*, 56(1), 1-43.
- [24] Campbell, J. Y., & Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3), 195-228.
- [25] Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1), 57-82.
- [26] Cao, J., & Han, B. (2013). Cross section of option returns and idiosyncratic stock volatility. *Journal of Financial Economics*, 108(1), 231-249.
- [27] Cetin, U., Jarrow, R., Protter, P., & Warachka, M. (2006). Pricing options in an extended Black Scholes economy with illiquidity: Theory and empirical evidence. *Review of Financial Studies*, 19(2), 493-529.

- [28] Cheng, M., & Madhavan, A. (2009). The dynamics of leveraged and inverse exchange-traded funds. *Journal Of Investment Management*, Fourth Quarter.
- [29] Chicago Board Options Exchange. (2009). Market maker program. Accessed from website at <https://www.cboe.org/publish/mmfaq/mmfaq.pdf> during November 2014.
- [30] Chicago Board Options Exchange, 2014, Product specifications. Accessed from website at <https://www.cboe.com/Products/EquityOptionSpecs.aspx> during August 2014.
- [31] Christie, W. G., & Schultz, P. H. (1994). Why do NASDAQ market makers avoid odd-eighth quotes?. *The Journal of Finance*, 49(5), 1813-1840.
- [32] Churm, R., Radia, A., Srinivasan, S., & Whisker, R. (2012). The Funding for Lending Scheme. *Bank of England Quarterly Bulletin*, 306-320.
- [33] Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies*, 21(4), 1533-1575.
- [34] Congressional Budget Office (CBO). (2009). The Troubled Asset Relief Program: Report on transactions through December 31, 2008.
- [35] Conrad, J. (1989). The price effect of option introduction. *The Journal of Finance*, 44(2), 487-498.
- [36] Cox, J. C., Ross, S. A., & Rubinstein, M. (1979). Option pricing: A simplified approach. *Journal of Financial Economics*, 7(3), 229-263.
- [37] Curdia, V., & Woodford, M. (2011). The central-bank balance sheet as an instrument of monetary policy. *Journal of Monetary Economics*, 58(1), 54-79.
- [38] Daniel, K., & Titman, S. (1997). Evidence on the characteristics of cross sectional variation in stock returns. *The Journal of Finance*, 52(1), 1-33.
- [39] Diamond, D. W., & Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *The Journal of Political Economy*, 91(3), 401-419.
- [40] De Ceuster, M. J., Dhaene, G., & Schatteman, T. (1998). On the hypothesis of psychological barriers in stock markets and Benford's Law. *Journal of Empirical Finance*, 5(3), 263-279.
- [41] Detemple, J., & Jorion, P. (1990). Option listing and stock returns: An empirical analysis. *Journal of Banking & Finance*, 14(4), 781-801.
- [42] Diether, K. B., Malloy, C. J., & Scherbina, A. (2002). Differences of opinion and the cross section of stock returns. *The Journal of Finance*, 57(5), 2113-2141.
- [43] Donaldson, R. G., & Kim, H. Y. (1993). Price barriers in the Dow Jones industrial average. *Journal of Financial and Quantitative Analysis*, 28(03), 313-330.

- [44] Duffie, D. (2010). Presidential address: Asset price dynamics with slow-moving capital. *The Journal of Finance*, 65(4), 1237-1267.
- [45] Easley, D., O'Hara, M., & Srinivas, P. S. (1998). Option volume and stock prices: Evidence on where informed traders trade. *The Journal of Finance*, 53(2), 431-465.
- [46] Frey, R., & Stremme, A. (1997). Market volatility and feedback effects from dynamic hedging. *Mathematical Finance*, 7(4), 351-374.
- [47] Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3-56.
- [48] Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy*, 81(3), 607-636.
- [49] Figlewski, S., & Webb, G. P. (1993). Options, short sales, and market completeness. *The Journal of Finance*, 48(2), 761-777.
- [50] Finnerty, J. E., & Park, H. Y. (1987). Stock index futures: Does the tail wag the dog?. *Financial Analysts Journal*, 43(2), 57-61.
- [51] Fischer, S. (2015 Feb 27). Conducting monetary policy with a large balance sheet. Federal Reserve Speech. Accessed from website at <http://www.federalreserve.gov/newsevents/speech/fischer20150227a.htm> during April 2015.
- [52] Frazzini, A., & Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, 111(1), 1-25.
- [53] Garleanu, N., & Pedersen, L. H. (2011). Margin-based asset pricing and deviations from the law of one price. *Review of Financial Studies*, 24(6), 1980-2022.
- [54] Garleanu, N., Pedersen, L. H., & Poteshman, A. M. (2009). Demand-based option pricing. *Review of Financial Studies*, 22(10), 4259-4299.
- [55] Ge, L., Lin, T. C., & Pearson, N. D. (2014). Why does the option stock volume ratio predict stock returns? Working Paper.
- [56] Geanakoplos, J. (2010). The leverage cycle. *NBER Macroeconomics Annual*, 24, 1-65.
- [57] Glosten, L. R., & Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1), 71-100.
- [58] Golez, B., & Jackwerth, J. C. (2012). Pinning in the S&P 500 futures. *Journal of Financial Economics*, 106(3), 566-585.
- [59] Green, T. C., & Figlewski, S. (1999). Market risk and model risk for a financial institution writing options. *The Journal of Finance*, 54(4), 1465-1499.
- [60] Greenwood, R. (2005). Short-and long-term demand curves for stocks: Theory and evidence on the dynamics of arbitrage. *Journal of Financial Economics*, 75(3), 607-649.

- [61] Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, 70(3), 393-408.
- [62] Harris, L. (1991). Stock price clustering and discreteness. *Review of Financial Studies*, 4(3), 389-415.
- [63] Harris, M., & Raviv, A. (1993). Differences of opinion make a horse race. *Review of Financial Studies*, 6(3), 473-506.
- [64] Harvey, C. R., Liu, Y., & Zhu, H. (2014). ... And the cross-section of expected returns. National Bureau of Economic Research Working Paper No. w20592.
- [65] He, Z., Khang, I. G., & Krishnamurthy, A. (2010). Balance sheet adjustments during the 2008 crisis. *IMF Economic Review*, 58(1), 118-156.
- [66] He, Z., & Krishnamurthy, A. (2013). Intermediary asset pricing. *American Economic Review*, 103(2), 732-770.
- [67] Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance*, 54(6), 2143-2184.
- [68] Hong, H., & Stein, J. C. (2007). Disagreement and the stock market. *Journal of Economic perspectives*, 21(2), 109-128.
- [69] Hu, J. (2014). Does option trading convey stock price information?. *Journal of Financial Economics*, 111(3), 625-645.
- [70] Jensen, M. C. & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4), 305-360.
- [71] Johnson, T. L., & So, E. C. (2012). The option to stock volume ratio and future returns. *Journal of Financial Economics*, 106(2), 262-286.
- [72] Kandel, E., & Pearson, N. D. (1995). Differential interpretation of public signals and trade in speculative markets. *Journal of Political Economy*, 103(4), 831-872.
- [73] Krishnamurthy, A. (2010). How debt markets have malfunctioned in the crisis. *Journal of Economic Perspectives*, 24(1), 3-28.
- [74] Krishnamurthy, A., & Vissing-Jorgensen, A. (2011). The effects of quantitative easing on interest rates: Channels and implications for policy. *Brookings Papers on Economic Activity*, 215-287.
- [75] Krishnamurthy, A., & Vissing-Jorgensen, A. (2013). The ins and outs of LSAPs. Working Paper.
- [76] Kubler, F., & Schmedders, K. (2012). Financial innovation and asset price volatility. *The American Economic Review*, 102(3), 147-151.
- [77] Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 53(6), 1315-1335.

- [78] Lakonishok, J., Lee, I., Pearson, N. D., & Poteshman, A. M. (2007). Option market activity. *Review of Financial Studies*, 20(3), 813-857.
- [79] Lehmann, B. N. (1990). Fads, martingales, and market efficiency. *Quarterly Journal of Economics*, 105 (1): 1-28.
- [80] Leland, H. E. (1985). Option pricing and replication with transactions costs. *The Journal of Finance*, 40(5), 1283-1301.
- [81] Ley, E., & Varian, H. R. (1994). Are there psychological barriers in the Dow-Jones index?. *Applied Financial Economics*, 4(3), 217-224.
- [82] Merrill, C. B., Nadauld, T., Stulz, R.M., & Sherlund, S. M. (2012). Did capital requirements and fair value accounting spark fire sales in distressed mortgage-backed securities? Ohio State University, Fisher College of Business Working Paper No. 2012-03-12.
- [83] Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, 4(1), 141-183.
- [84] Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. *The Journal of Finance*, 32(4), 1151-1168.
- [85] Mitchell, M., Pedersen, L. H., & Pulvino., T. (2007). Slow moving capital. *American Economic Review*, 97(2), 215–20.
- [86] Mitchell, M., & Pulvino, T. (2012). Arbitrage crashes and the speed of capital. *Journal of Financial Economics*, 104(3), 469-490.
- [87] Nagel, S. (2012). Evaporating liquidity. *Review of Financial Studies*, 25(7), 2005-2039.
- [88] Newey, W. K., & West, K. D. (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703-708.
- [89] Ni, S. X., Pearson, N. D., & Poteshman, A. M. (2005). Stock price clustering on option expiration dates. *Journal of Financial Economics*, 78(1), 49-87.
- [90] Myers, S. C. (1977). Determinants of corporate borrowing. *Journal of Financial Economics*, 5(2), 147-175.
- [91] Pan, J., & Poteshman, A. M. (2006). The information in option volume for future stock prices. *Review of Financial Studies*, 19(3), 871-908.
- [92] Pastor, L., & Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, 111(3), 642-685.
- [93] Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies*, 22(1), 435-480.
- [94] Platen, E., & Schweizer, M. (1998). On feedback effects from hedging derivatives. *Mathematical Finance*, 8(1), 67-84.

- [95] Pomorski, L. (2008). Stock performance after crossing a round-number price. Working Paper.
- [96] Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance*, 39(4), 1127-1139.
- [97] Roll, R., Schwartz, E., & Subrahmanyam, A. (2010). O/S: The relative trading activity in options and stock. *Journal of Financial Economics*, 96(1), 1-17.
- [98] Scholes, M., & Williams, J. (1977). Estimating betas from nonsynchronous data. *Journal of Financial Economics*, 5(3), 309-327.
- [99] Schonbucher, P. J., & Wilmott, P. (2000). The feedback effect of hedging in illiquid markets. *SIAM Journal on Applied Mathematics*, 61(1), 232-272.
- [100] Shleifer, A. (1986). Do demand curves for stocks slope down?. *The Journal of Finance*, 41(3), 579-590.
- [101] Shleifer, A., & Vishny, R. W. (1992). Liquidation values and debt capacity: A market equilibrium approach. *The Journal of Finance*, 47(4), 1343-1366.
- [102] Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance*, 52(1), 35-55.
- [103] Shleifer, A., & Vishny, R. W. (2010). Asset fire sales and credit easing. *American Economic Review Papers and Proceedings*, 100(2), 46-50.
- [104] Sorescu, S. M. (2000). The effect of options on stock prices: 1973 to 1995. *The Journal of Finance*, 55(1), 487-514.
- [105] Stiglitz, J. E., & Weiss, A. (1981). Credit rationing in markets with imperfect information. *The American Economic Review*, 71(3), 393-410.
- [106] Subrahmanyam, M. G., Tang, D. Y., & Wang, S. Q. (2014). Does the tail wag the dog?: The effect of credit default swaps on credit risk. *Review of Financial Studies*, 27(10), 2927-2960.
- [107] Wooldridge, J. M. (2001). *Econometric analysis of cross section and panel data*. MIT press.
- [108] Wurgler, J., & Zhuravskaya, E. (2002). Does arbitrage flatten demand curves for stocks?. *The Journal of Business*, 75(4), 583-608.
- [109] Yang, D. C. (2015). Disagreement and the option stock volume ratio. In Preparation.

Appendix A

Proofs to Chapter 2

Most of the proofs are computing the derivative and recognizing different forms of the regularity condition. The regularity condition in Equation 2.5 ensures that the shorting constraint binds. In the proofs, I apply several different forms of the regularity condition, which are all equivalent. I list them here:

$$\begin{aligned} p_x^{*,bmark} - c &> \mu - h \\ \Leftrightarrow h &> c + \frac{Q_x}{\tau} \\ \Leftrightarrow h\tau - c\tau - Q_x &> 0 \\ \Leftrightarrow 4h^2\tau - 4hQ_x - 4ch\tau &= 4h(h\tau - Q_x - c\tau) > 0 \\ \Leftrightarrow 2h - c &> h + \frac{Q_x}{\tau} > 0 \\ \Leftrightarrow c - 2h &< -h - \frac{Q_x}{\tau} < 0 \end{aligned}$$

It is also convenient to solve out Q_z^* in terms of the underlying parameters of the model. Combining Equation 2.4 and Equation 2.6, we get:

$$Q_z^* = \frac{1}{4hn} \cdot \tau \left(\frac{c}{2} + h \left(\frac{2Q_x}{2h\tau - c\tau} - 1 \right) \right) \quad (\text{A.1})$$

Proof of Proposition 4: Applying Equation 2.3 and Equation 2.4, we get

$$\begin{aligned}
p_x^* - p_x^{*,bmark} &= c\left(\frac{1}{2} - \frac{1}{2h-c} \frac{Q_x}{\tau}\right) \\
&= c\left(\frac{2h-c-2\frac{Q_x}{\tau}}{2(2h-c)\tau}\right) \\
&> 0
\end{aligned}$$

The denominator is positive since $2h - c > 0$ by the regularity condition. The numerator is also positive since $2h - c > 2h - 2c > 2\frac{Q_x}{\tau}$, where the last inequality follows from the regularity condition.

Proof of Proposition 5: Using Equation A.1 and taking the derivative with respect to the belief heterogeneity c , we get:

$$\frac{\partial Q_z^*}{\partial c} = \frac{-1}{8h(2h-c)^3} (c^2\tau + 4h^2\tau - 4hQ_x - 4ch\tau)(c^2\tau + 4h^2\tau + 4hQ_x - 4ch\tau)$$

By the regularity condition, $2h - c > 0$. Also $4h^2\tau - 4hQ_x - 4ch\tau > 0$, by the regularity condition. This implies that $4h^2\tau + 4hQ_x - 4ch\tau > 4h^2\tau - 4hQ_x - 4ch\tau > 0$. Therefore, the derivative is negative.

Proof of Proposition 6: Using Equation A.1 and taking the derivative with respect to the belief heterogeneity h , we get:

$$\frac{\partial Q_z^*}{\partial h} = \frac{c^2\tau + 4h^2\tau - 4h(Q_x + c\tau)}{16h^2(2h-c)^3\tau^2} \cdot \frac{(c+2h)(c^2\tau + 4h^2\tau + 4h(Q_x - c\tau))}{n}$$

By the regularity condition, $2h - c > 0$. Next, $4h^2\tau - 4h(Q_x + c\tau) = 4h(h\tau - Q_x - c\tau) > 0$ by the regularity condition. Finally, $4h^2\tau + 4h(Q_x - c\tau) = 4h(h\tau + Q_x - c\tau) > 4h(h\tau - Q_x - c\tau) > 0$ where the last step follows from the regularity condition. Therefore, $\frac{\partial Q_z^*}{\partial h} > 0$.

Proof of Proposition 7: Using Equation 2.4 and taking the derivative with respect to the belief heterogeneity h , we get:

$$\frac{\partial p_x^*}{\partial h} = \frac{2cQ_x}{(c-2h)^2\tau} > 0 \quad (\text{A.2})$$

Hence we immediately have that the derivative is positive.

Proof of Proposition 8: Using Equation A.1 and taking the derivative with respect to

the risk tolerance τ , we get:

$$\frac{\partial Q_z^*}{\partial \tau} = \frac{1}{n} \frac{Q_x}{2(c-2h)^2\tau^3} \cdot (c^2\tau + 4h^2\tau - 4h(Q_x + c\tau)) > 0$$

The denominator is positive. Furthermore, $4h^2\tau - 4h(Q_x + c\tau) = 4h(h\tau - (Q_x + c\tau)) > 0$ by the regularity condition.

Next, using Equation 2.4 and taking the derivative with respect to the risk tolerance τ , we get:

$$\frac{\partial p_x^*}{\partial \tau} = \frac{2hQ}{(2h-c)\tau^2} > 0 \quad (\text{A.3})$$

since $2h - c > 0$ by the regularity condition.

Proof of Proposition 9: Using Equation 2.4 and taking the derivative with respect to the quantity of the underlying stock Q_x , we get:

$$\frac{\partial p_x^*}{\partial Q_x} = -\frac{2h}{(2h-c)\tau} < 0 \quad (\text{A.4})$$

since $2h - c > 0$ by the regularity condition.

Using Equation A.1 and taking the derivative with respect to the quantity of the underlying stock Q_x , we get:

$$\frac{\partial Q_z^*}{\partial Q_x} = -\frac{1}{n} \frac{c^2\tau + 4h(h\tau - (Q_x + c\tau))}{2(2h-c)^2\tau^2} < 0$$

since $2h - c > 0$ and $h\tau - (Q_x + c\tau) > 0$ by the regularity condition.

Proof of Proposition 10: Using Equation A.3, I take another derivative with respect to h to get:

$$\frac{\partial p_x^*}{\partial \tau \partial h} = \frac{2cQ_x}{(c-2h)^2\tau} < 0$$

However, the derivative $\frac{\partial p_x^*}{\partial \tau}$ overall is still positive, by Proposition 8,

Then, using Equation A.4, I also take another derivative with respect to h to get:

$$\frac{\partial p_x^*}{\partial Q_x \partial h} = \frac{2c}{(c-2h)^2\tau} > 0$$

However, the derivative $\frac{\partial p_x^*}{\partial Q_x}$ overall is still negative, by Proposition 9.

Appendix B

Proofs to Chapter 3

Proof to Proposition 11: At the interior solution, $p_x = p_y = p$. Therefore, the system of equations reduces to:

$$I^* + pQ = N^B + (\tau_x + \tau_y) \quad (\text{B.1})$$

$$f'(I^*) = \frac{1}{p} \quad (\text{B.2})$$

Notice that only the total size of the intervention $\tau_x + \tau_y$ matters for the equilibrium.

Proof to Proposition 12: The equilibrium of the equity injection is composed of the first order conditions:

$$\begin{aligned} \frac{1}{p_x} &= \frac{1}{p_y} \\ f'(I) &= \frac{1}{p_x} \end{aligned}$$

and the market clearing conditions:

$$\begin{aligned} \frac{I}{p_x} &= x^B \\ Q &= y^B \end{aligned}$$

and the budget constraint $N^B + \tau_x + \tau_y = p_x x^B + p_y y^B$. At the interior solution $p_x = p_y$ and this system of equations reduces to

$$I^* + pQ = N^B + \tau_x + \tau_y \quad (\text{B.3})$$

$$f'(I^*) = \frac{1}{p} \quad (\text{B.4})$$

which is the same system of equations as in the proof to Proposition 11. Therefore, $p^{*EqtyInjection} = p^{*Direct}$.

Proof to Proposition 13: Using Equation 3.15, solve for I^* in terms of p_x . I define a function $g(p_x) = I^*$. Then, use Equations 3.13 and 3.14, solve for p_y and h_y in terms of the parameters in the model.

$$\begin{aligned} p_y &= \frac{(\tau_x + \tau_y)(1 - p_x) + h_x p_x Q + (h_x - 1 + p_x - h_x p_x)g(p_x)}{(1 + (h_x - 1)p_x)Q} \\ h_y &= \frac{h_x p_x (-\tau_x - \tau_y + Q + (1 - h_x)g(p_x))}{h_x p_x Q + (\tau_x + \tau_y)(1 - p_x) + (h_x - 1 + p_x - h_x p_x)g(p_x)} \end{aligned}$$

The market clearing condition is

$$g(p_x) + p_y Q = N^B + (p_x - h_x p_x) \frac{g(p_x)}{p_x} + (p_y - h_y p_y) Q$$

Plug in p_y and h_y to get:

$$\frac{(\tau_x + \tau_y)(1 - p_x) + h_x p_x Q + h_x g(p_x)}{1 + (h_x - 1)p_x} = \tau_x + \tau_y + N^B$$

Take the implicit derivative and collect terms to get

$$\begin{aligned} \frac{\partial p_x}{\partial h_x} &= \frac{(\tau_x + \tau_y + N^B - Q - \frac{g(p_x)}{p_x})p_x}{N^B - h_x(\tau_x + \tau_y + N^B - Q - g'(p_x))} \\ &< 0 \end{aligned}$$

The parameter restriction is $N^B + \tau_x + \tau_y < I + Q$. Also, since $p_x < 1$, we have $g(p_x)/p_x > g'(p_x)$. Therefore, the numerator is negative. The denominator is positive since there are diminishing margin returns to the production function. That is, $f'(I) = 1/p_x$ therefore we can take the implicit derivative $g'(p_x) = -1/(p_x^2 f''(I))$.

Proof to Proposition 14: By the bank's first order condition, we have

$$\frac{1 - (p_x - h_x p_x)}{h_x p_x} = \frac{1 - (p_y - h_y p_y)}{h_y p_y}$$

When $h_x = h_y$, we get $p_x = p_y$.

Market clearing conditions give us

$$I^* + p_y Q = N^B + \tau_x + \tau_y$$

where $f'(I^*) = \frac{1}{p_x}$. Hence, when $p = p_x = p_y$, then we get $I^* + pQ = N^B + \tau_x + \tau_y$ where $f'(I^*) = \frac{1}{p}$. This equation is identical to Equation B.1, the equation that determines $p^{*Direct}$.